

10 tout ou partie du gène DP428, objet de la présente invention conduirait probablement à une meilleure protection contre la tuberculose. Tout ou partie du gène DP428, ou tout polynucléotide selon l'invention, peut être
15 facilement inséré dans les plasmides vecteurs V1J (Montgomery et al, 1993), pcDNA3 (Invitrogen, R & D Systems) ou pcDNA1/Neo (Invitrogen) qui possèdent les caractéristiques nécessaires pour une utilisation vaccinale.

20 L'invention vise ainsi un vaccin, caractérisée en ce qu'il comprend un ou plusieurs polypeptides selon l'invention et/ou un ou plusieurs polypeptides hybrides selon l'invention tels que précédemment définis en
25 association avec un véhicule pharmaceutiquement compatible et, le cas échéant, un ou plusieurs adjuvants de l'immunité appropriés.

L'invention vise aussi une composition vaccinale
30 destinée à l'immunisation de l'homme ou l'animal à l'encontre d'une infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, caractérisée en ce qu'elle comprend un ou plusieurs polypeptides hybrides tels que précédemment définis en association avec un véhicule
35 pharmaceutiquement compatible et, le cas échéant, un ou plusieurs adjuvants de l'immunité.

Avantageusement, dans le cas d'une protéine hybride entre un polypeptide selon l'invention et l'antigène de
40 surface de l'hépatite B, la composition vaccinale sera administrée, chez l'homme, à raison de 0,1 à 1 µg de protéine hybride purifiée par kilogramme du poids du patient, de préférence 0,2 à 0,5 µg/kg de poids du patient, pour une dose destinée à une administration donnée. Dans le
45 cas de patients atteints de troubles du système immunitaire, en particulier les patients immunodéprimés, chaque dose injectée contiendra préférentiellement la

moitié de la quantité pondérale de la protéine hybride contenue dans une dose destinée à un patient n'étant pas affecté de troubles du système immunitaire.

5 De préférence, la composition vaccinale sera administrée à plusieurs reprises, de manière étalée dans le temps, par voie intradermique ou sous-cutanée. A titre d'exemple, trois doses telles que définies ci-dessus seront respectivement administrées au patient au temps t_0 , au
10 temps $t_0 + 1$ mois et au temps $t_0 + 1$ an.

Alternativement, trois doses seront respectivement administrées au patient au temps t_0 , au temps $t_0 + 1$ mois et au temps $t_0 + 6$ mois.

15 Chez la souris, chez laquelle une dose pondérale de la composition vaccinale comparable à la dose utilisée chez l'homme est administrée, la réaction anticorps est testée par prélèvement du sérum suivi d'une étude de la formation d'un complexe entre les anticorps présents dans le sérum et
20 l'antigène de la composition vaccinale, selon les techniques usuelles.

L'invention concerne également une composition immunogène caractérisée en ce qu'elle comprend un
25 polynucléotide ou un vecteur d'expression selon l'invention, en association avec un véhicule permettant son administration à l'homme ou l'animal.

L'invention a encore pour objet un vaccin destiné à
30 l'immunisation à l'encontre d'une infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, caractérisé en ce qu'il comprend un polynucléotide ou un vecteur d'expression selon l'invention, en association avec un véhicule pharmaceutiquement acceptable.

35

De telles compositions immunogènes ou vaccinales sont notamment décrites dans la demande internationale N° WO

90/11092 (Vical Inc.) et également dans la demande internationale N° WO 95/11307 (Institut Pasteur).

Le polynucléotide constitutif de la composition immunogène ou de la composition vaccinale selon l'invention peut être injecté à l'hôte après avoir été couplé à des composés qui favorisent la pénétration de ce polynucléotide à l'intérieur de la cellule ou son transport jusqu'au noyau cellulaire. Les conjugués résultants peuvent être encapsulés dans des microparticules polymères, comme décrit dans la demande internationale N° WO 94/27238 (medisorb Technologies International).

Selon un autre mode de réalisation de la composition immunogène et/ou vaccinale selon l'invention, le polynucléotide, de préférence un ADN, est complexé avec du DEAE-dextran (Pagano et al., 1967) ou avec des protéines nucléaires (Kaneda et al., 1989), avec des lipides (Feigner et al., 1987) ou encore encapsulés dans des liposomes (Fraley et al., 1980).

Selon encore un autre mode de réalisation avantageux de la composition immunogène et/ou vaccinale selon l'invention, le polynucléotide selon l'invention peut être introduit sous la forme d'un gel facilitant sa transfection dans les cellules. Une telle composition sous forme de gel peut être un complexe de poly-L-lysine et de lactose, comme décrit par Midoux en 1993, ou encore le Poloxamer 407™, comme décrit par Pastore en 1994. Le polynucléotide ou le vecteur selon l'invention peuvent aussi être en suspension dans une solution tampon ou être associés à des liposomes.

Avantageusement, un tel vaccin sera préparé conformément à la technique décrite par Taceon et al. ou Huygen et al. en 1996 ou encore conformément à la technique décrite par Davis et al. dans la demande internationale N° WO 95/11307 (Whalen et al.).

Un tel vaccin sera avantageusement préparé sous la forme d'une composition contenant un vecteur selon l'invention, placée sous le contrôle d'éléments de régulation permettant son expression chez l'homme ou l'animal.

Pour réaliser un tel vaccin, le polynucléotide selon l'invention est tout d'abord sous-cloné dans un vecteur d'expression approprié, plus particulièrement un vecteur d'expression contenant des signaux de régulation et d'expression reconnus par les enzymes des cellules eucaryotes et contenant également une origine de répllication active chez les procaryotes, par exemple chez *E. coli*, qui permet son amplification préalable. Puis le plasmide recombinant purifié obtenu est injecté à l'hôte, par exemple par voie intramusculaire.

On pourra par exemple utiliser, en tant que vecteur d'expression in vivo de l'antigène d'intérêt, le plasmide pcDNA3 ou le plasmide pcDNA1/neo, tous les deux commercialisés par Invitrogen (R&D Systems, Abingdon, Royaume-Uni). On peut aussi utiliser le plasmide V1Jns.tPA, décrit par Shiver et al. en 1995.

Un tel vaccin comprendra avantageusement, outre le vecteur recombinant, une solution saline, par exemple une solution de chlorure de sodium.

Une composition vaccinale telle que définie ci-dessus sera par exemple administrée par voie parentérale ou par voie intramusculaire.

La présente invention concerne également un vaccin caractérisé en ce qu'il contient une ou plusieurs séquences nucléotidiques selon l'invention et/ou un ou plusieurs polynucléotides tel que mentionné ci-dessus en association avec un véhicule pharmaceutiquement compatible et, le cas

échéant, un ou plusieurs adjuvants de l'immunité appropriés.

Un autre aspect porte sur une méthode de criblage de
5 molécules capables d'inhiber la croissance de mycobactéries
ou le maintien de mycobactéries dans un hôte, caractérisée
en ce que lesdites molécules bloquent la synthèse ou la
fonction des polypeptides codés par une séquence
nucléotidique selon l'invention ou par un polynucléotide
10 tel que décrit supra.

Dans ladite méthode de criblage, les molécules peuvent
être des anti-messagers ou peuvent induire la synthèse
d'anti-messagers.

15

La présente invention vise également des molécules
capables d'inhiber la croissance de mycobactéries ou le
maintien de mycobactéries dans un hôte, caractérisées en ce
que lesdites molécules sont synthétisées d'après la
20 structure des polypeptides codés par une séquence
nucléotidique selon l'invention ou par un polynucléotide
tel que décrit supra.

D'autres caractéristiques et avantages de l'invention
25 apparaissent dans les exemples et les figures suivants :

FIGURES

30 La série de Figures 1 :

La série de Figures 1 illustre la série de séquences
nucléotidiques SEQ ID N°1 correspondant à l'insert du
vecteur pDP428 (déposé à la CNOM sous le N° I-1818) et
35 la série de séquences d'acides aminés SEQ ID N°1 des

polypeptides codés par la série des séquences nucléotidiques SEQ ID N°1.

Figure 2 :

5

Illustre la séquence nucléotidique SEQ ID N°2 correspondant à la région incluant le gène codant pour le polypeptide DP428 (région soulignée). Sur cette figure ont été pris en compte à la fois les codons ATG et GTG d'initiation de la traduction. La figure fait apparaître que le polypeptide DP428 fait probablement partie d'un opéron comprenant au moins trois gènes. La région doublement encadrée inclut probablement les régions promotrices.

10

La région simplement encadrée correspond au motif LPISG rappelant le motif LPXTG décrit chez les bactéries à Gram positifs comme permettant l'ancrage aux peptidoglycannes.

15

20 La série de Figures 3 :

La série de Figures 3 représente la série de séquences nucléotidiques SEQ ID N°3 correspondant à l'insert du vecteur p6D7 (déposé à la CNCM sous le N° I-1814).

25

La série de Figures 4 :

La série de Figures 4 représente la série de séquences nucléotidiques SEQ ID N°4 correspondant à l'insert du vecteur p5A3 (déposé à la CNCM sous le N° I-1815).

30

La série de Figures 5 :

La série de Figures 5 représente la série de séquences nucléotidiques SEQ ID N°5 correspondant à l'insert du vecteur p5F6 (déposé à la CNCM sous le N° I-1816).

35

La série de Figures 6 :

La série de Figures 6 représente la série de séquences
nucléotidiques SEQ ID N°6 correspondant à l'insert du
vecteur p2A29 (déposé à la CNCM sous le N° I-1817).

La série de Figures 7 :

La série de Figures 7 représente la série de séquences
nucléotidiques SEQ ID N°7 correspondant à l'insert du
vecteur p5B5 (déposé à la CNCM sous le N° I-1819).

La série de Figures 8 :

La série de Figures 8 représente série de séquences
nucléotidiques SEQ ID N°8 correspondant à l'insert du
vecteur p1C7 (déposé à la CNCM sous le N° I-1820).

La série de Figures 9 :

La série de Figures 9 représente la série de séquences
nucléotidiques SEQ ID N°9 correspondant à l'insert du
vecteur p2D7 (déposé à la CNCM sous le N° I-1821).

La série de Figures 10 :

La série de Figures 10 représente la série de séquences
nucléotidiques SEQ ID N°10 correspondant à l'insert du
vecteur p1B7 (déposé à la CNCM sous le N° I-1843).

La série de Figures 11 :

La série de Figures 11 représente la série de séquences
nucléotidiques SEQ ID N°11.

La série de Figures 12 :

La série de Figures 12 représente la série de séquences nucléotidiques SEQ ID N°12.

La série de Figures 13 :

5

La série de Figures 13 représente la série de séquences nucléotidiques SEQ ID N°13.

La série de Figures 14 :

10

La série de Figures 14 représente la série de séquences nucléotidiques SEQ ID N°14 correspondant à l'insert du vecteur pSB5 (déposé à la CNCM sous le N° I-1819).

15

La série de Figures 15 :

20 La série de Figures 15 représente la série de séquences nucléotidiques SEQ ID N°15.

La série de Figures 16 :

25 La série de Figures 16 représente la série de séquences nucléotidiques SEQ ID N°16.

La série de Figures 17 :

30 La série de Figures 17 représente la série de séquences nucléotidiques SEQ ID N°17.

La série de Figures 18 :

35 La série de Figures 18 représente la série de séquences nucléotidiques SEQ ID N°18.

La série de Figures 19 :

La série de Figures 19 représente la série de séquences nucléotidiques SEQ ID N°19.

9 La série de Figures 20 :

La série de Figures 20 représente la série de séquences nucléotidiques SEQ ID N°20 correspondant à l'insert du vecteur p2A29 (déposé à la CNCM sous le N° I-1817).

10

La série de Figures 21 :

La série de Figures 21 représente la série de séquences nucléotidiques SEQ ID N°21.

15

La série de Figures 22 :

La série de Figures 22 représente la série de séquences nucléotidiques SEQ ID N°22.

20

La série de Figures 23 :

La série de Figures 23 représente la série de séquences nucléotidiques SEQ ID N°23.

25

La série de Figures 24 :

La série de Figures 24 représente la série de séquences nucléotidiques SEQ ID N°24.

30

Figures 25 et 26 :

Les figures 25 et 26 illustrent respectivement les séquences SEQ ID N°25 et SEQ ID N°26 représentant un couple d'amorces utilisées pour amplifier spécifiquement par PCR la région correspondant aux nucléotides 964 à 1234 inclus dans la séquence SEQ ID N°1.

35

La série de Figures 27 :

La série de Figures 27 représente la série de séquences
nucléotidiques SEQ ID N°27 correspondant à l'insert du
vecteur p5A3.

Figure 28 :

La séquence d'acides aminés telle que définie dans la
figure 28 représente la séquence d'acides aminés SEQ ID
N°28 correspondant au polypeptide DP428.

Figure 29 :

La figure 29 représente la séquence nucléotidique SEQ ID N°
29 du gène complet codant pour la protéine M1C25.

Figure 30 :

La figure 30 représente la séquence d'acides aminés SEQ ID
N° 30 de la protéine M1C25.

La série de Figures 31 :

La série de Figures 31 représente la série de séquences
nucléotidiques SEQ ID N°31.

La série de Figures 32 :

La série de Figures 32 représente la série de séquences
nucléotidiques SEQ ID N°32.

La série de Figures 33 :

La série de Figures 33 représente la série de séquences nucléotidiques SEQ ID N°33.

La série de Figures 34 :

5

La série de Figures 32 représente la série de séquences nucléotidiques SEQ ID N°34.

La série de Figures 35 :

10

La série de Figures 35 représente la série de séquences nucléotidiques SEQ ID N°35.

15

La série de Figures 36 :

La série de Figures 36 représente la série de séquences nucléotidiques SEQ ID N°36.

20

La série de Figures 37 :

La série de Figures 37 représente la série de séquences nucléotidiques SEQ ID N°37.

25

La série de Figures 38 :

La série de Figures 38 représente la série de séquences nucléotidiques SEQ ID N°38.

30

La série de Figures 39 :

La série de Figures 39 représente la série de séquences nucléotidiques SEQ ID N°39.

35

La série de Figures 40 :

La série de Figures 40 représente la série de séquences nucléotidiques SEQ ID N°40.

La série de Figures 41 :

5

La série de Figures 41 représente la série de séquences nucléotidiques SEQ ID N°41 correspondant à l'insert du vecteur p2D7 (déposé à la CNCM sous le N°I-1821).

10 La série de Figures 42 :

La série de Figures 42 représente la série de séquences nucléotidiques SEQ ID N°42.

15

La série de Figures 43 :

20 La série de Figures 43 représente la série de séquences nucléotidiques SEQ ID N°43.

La série de Figures 44 :

25 La série de Figures 44 représente la série de séquences nucléotidiques SEQ ID N°44.

La série de Figures 45 :

30 La série de Figures 45 représente la série de séquences nucléotidiques SEQ ID N°45.

La série de Figures 46 :

35 La série de Figures 46 représente la série de séquences nucléotidiques SEQ ID N°46.

La série de Figures 47 :

La série de Figures 47 représente la série de séquences nucléotidiques SEQ ID N°47.

5 La série de Figures 48 :

La série de Figures 48 représente la série de séquences nucléotidiques SEQ ID N°48.

10 La série de Figures 49 :

La série de Figures 49 représente la série de séquences nucléotidiques SEQ ID N°49.

15

La série de Figures 50 :

La série de Figures 50 représente la série de séquences nucléotidiques SEQ ID N°50.

20

Figure 51 :

A. la construction pJVED: Plasmid navette(pouvant se multiplier chez les mycobactéries ainsi que chez
25 E.coli). avec un gène de résistance à la kanamycine (issu de Tn903) comme marqueur de sélection. Le gène *phoA* tronqué(Δ *phoA*) et le gène *luc* forment un opéron synthétique.

B. Séquence de la jonction entre *phoA* et *luc*.

30

Figure 51 :

Hybridation génomique (Southern blot) de l'ADN génomique de différentes espèces mycobactériennes à
35 l'aide d'une sonde oligonucléotidique dont la séquence est la séquence comprise entre le nucléotide en

position nt 964 (extrémité 5' de la sonde) et le nucléotide en position nt 1234 (extrémité 3' de la sonde), extrémités incluses, de la séquence SEQ ID N°1.

5 Figures 53 et 54 :

Activités Luc et PhcA de *M. smegmatis* recombinant contenant le pJVED avec différents fragments nucléotidiques comme décrites en exemple. Les figures 52
10 et 53 représentent les résultats obtenus pour deux expériences distinctes réalisées dans les mêmes conditions.

15

Figure 55 :

Représentation de l'hydrophobicité (Kyte et Doolittle) de la séquence codante du polypeptide DP428 avec sa
20 représentation schématique. Le motif LPISG précède immédiatement la région C-terminale hydrophobe. La séquence se termine par deux arginines.

Figure 56 :

25 Représentation de l'hydrophobicité (Kyte et Doolittle) de la séquence du polypeptide M1C25 de séquence d'acides aminés SEQ ID N° 30.

Figure 57 :

30

A- Gel d'acrylamide (12%) en condition dénaturante d'un extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 sans et après 4 heures d'induction par l'IPTG, coloré au bleu de
35 Comassie.

ligne 1: Marqueur de masse molaire (Prestained SDS-PAGE Standards High Range BIO-RAD®).

5 ligne 2: Extrait bactérien obtenu par sonication de bactéries E. coli M15 contenant le plasmide pM1C25 sans induction par l'IPTG.

10 ligne 3: Extrait bactérien obtenu par sonication de bactéries E. coli M15 contenant le plasmide pM1C25 après 4 heures d'induction par l'IPTG.

ligne 4: Marqueur de masse molaire (Prestained SDS-PAGE Standards Low Range BIO-RAD®).

15

B- Western blot d'un gel semblable gel (acrylamide 12%) révélé grâce à l'anticorps penta-His commercialisé par la société Quiagen.

20 ligne 1: représentation du marqueur de masse molaire (Prestained SDS-PAGE Standards High Range BIO-RAD®).

25 ligne 2: extrait bactérien obtenu par sonication de bactéries E. coli M15 contenant le plasmide pM1C25 sans induction par l'IPTG.

30 ligne 3: extrait bactérien obtenu par sonication de bactéries E. coli M15 contenant le plasmide pM1C25 après 4 heures d'induction par l'IPTG.

ligne 4: représentation du marqueur de masse molaire (Prestained SDS-PAGE Standards Low Range BIO-RAD®)

35 La bande présente très majoritairement dans les lignes correspondant aux bactéries induites par l'IPTG par rapport à celles non induites par l'IPTG, comprise entre 34200 et

28400 daltons, correspond à l'expression de l'insert M1C25 cloné dans le vecteur pQE-60 (Qiagen®).

5 En ce qui concerne les légendes des autres figures qui sont numérotées par un caractère alphanumérique, chacune de ces autres figures représente la séquence nucléotidique et la séquence d'acides aminés de séquence SEQ ID dont la numérotation est identique au caractère alphanumérique de
10 chacune desdites figures.

Les numérotations alphanumériques des figures représentant les SEQ ID comportant un nombre suivi d'une lettre ont les significations suivantes :

- les numérotations alphanumériques présentant le même
15 nombre concernant une même famille de séquence rattachées à la séquence de référence SEQ ID dont la numérotation présente ce même nombre et la lettre A ;
- les lettres A, B et C pour une même famille de séquences distinguent les trois phases de lecture possibles de la
20 séquence nucléotidique SEQ ID de référence (A) ;
- les lettres indexées par un prime (') signifient que la séquence correspond à un fragment de la séquence SEQ ID de référence (A) ;
- la lettre D signifie que la séquence correspond à la
25 séquence du gène prédit par Cole et al., 1998 ;
- la lettre F signifie que la séquence correspond à la phase ouverte de lecture (ORF pour "Open Reading Frame") contenant la séquence "D" correspondante d'après Cole et al., 1998 ;
- 30 - la lettre G signifie que la séquence est une séquence prédite par Cole et al., 1998, et présentant une homologie de plus de 70% avec la séquence SEQ ID de référence (A) ;
- la lettre H signifie que la séquence correspond à la phase ouverte de lecture contenant la séquence "G"
35 correspondante d'après Cole et al., 1998 ;
- la lettre R signifie que la séquence correspond à une séquence prédite par Cole et al., 1998, en amont de la

séquence "D" correspondante et pouvant être en phase avec la séquence "B" en raison d'erreurs de séquençage possibles ;

- la lettre P signifie que la séquence correspond à la phase ouverte de lecture contenant la séquence "R" correspondante ;
- la lettre Q signifie que la séquence correspond à une séquence contenant les séquences "P" et "p" correspondantes.

En ce qui concerne la famille de séquences SEQ ID N° 4, l'insert précédent *phoA* contient deux fragments non contigus sur le génome, SEQ ID 4J et SEQ ID 4A, et donc issus d'un clonage multiple permettant l'expression et l'exportation de *phoA*. Ces deux fragments non contigus, les gènes et les phases ouvertes de lecture qui les contiennent d'après Cole et al., 1998, sont importants pour l'exportation d'un polypeptide antigène ;

- les lettres J, K et L distinguant les trois phases de lecture possibles de la séquence nucléotidique "J" correspondante ;
- la lettre M signifie que la séquence correspond à la séquence prédite par Cole et al., 1998, et contenant la séquence SEQ ID N° 4J ;
- la lettre N signifie que la séquence correspond à la phase ouverte de lecture contenant la séquence SEQ ID N° 4M.

En ce qui concerne la famille de séquences SEQ ID N° 45, la lettre Z signifie que la séquence correspond à la séquence d'un fragment cloné fusionné avec *phoA*.

Enfin, en ce qui concerne la famille de séquence SEQ ID N° 41, la lettre S signifie que la séquence correspond à une séquence prédite par Cole et al., 1998 et pouvant être dans la même phase de lecture que la séquence "D" correspondante, la lettre T signifiant que la séquence correspondante contient les séquences "F" et "G" correspondantes.

EXEMPLES

Matériel et méthodes

5 Cultures bactériennes, plasmides et milieux de cultures

E. coli a été cultivé sur milieu liquide ou solide Luria-Bertani (LB). *M. smegmatis* a été cultivé sur milieu liquide Middlebrook 7H9 (Difco) additionné de dextrose albumine (ADC), 0,2 % de glycérol et 0,05 % de Tween, ou sur milieu
10 solide L. Si nécessaire, l'antibiotique kanamycine a été rajouté à une concentration de 20 $\mu\text{g/ml}^{-1}$. Les clones bactériens présentant une activité PhoA ont été détectés sur de l'agar LB contenant du 5-bromo-4-chloro-3-indolyile
15 phosphate (X-P, à 40 $\mu\text{g/ml}^{-1}$).

Manipulation d'ADN et séquençage

Les manipulations d'ADN et les analyses par
20 Southern blot ont été effectuées en utilisant les techniques standard (Sambrook et al., 1989). Les séquences d'ADN double brin ont été déterminées avec un kit de séquençage Taq Dye Deoxy Terminator Cycle (Applied Biosystems), dans un Système 9600 GeneAmp PCR (Perkin-
25 Elmer), et après migration sur un système d'analyse ADN modèle 373 (Applied Biosystems).

Constructions des plasmides

30 Le plasmide pJVEDa a été construit à partir de pLA71, plasmide de transfert comportant le gène *phoA* tronqué et placé en phase avec *BlaF*. pLA71 a été coupé avec les enzymes de restriction *KpnI* et *NotI*, retirant ainsi *phoA* sans toucher le promoteur de *BlaF*. Le gène *luc* codant
35 pour la luciférase de luciole a été amplifié à partir de

pGEM-luc et un site de liaison du ribosome a été rajouté. *phoA* a été amplifié à partir de pJEM11. Les fragments amplifiés ont été coupés avec *Pst*I et ligaturés ensemble. Les oligodéoxynucléotides utilisés sont les suivants :

- 5 pPV.luc.Fw : 5'GACTGCTGCACAAGCAGAAGATCCAAATGG3'
 luc.Bw : 5'GACTAGCGGCGCGGAATTGCTCGACCTCCGAGG3'
 pJEM.phoA.Fw : 5'CCGCGGATCCCGGATACGTAC3'
 phoA.Bw : 5'GACTGCTGCAGTTTATTTCAGCCCCAGAGCG3'.

Le fragment ainsi obtenu a été réamplifié en utilisant les oligonucléotides complémentaires de ses extrémités, coupé avec *Kpn*I et *Not*I, et intégré dans pLA71 coupé avec les mêmes enzymes. La construction résultante a été électroporée dans *E. coli* DH5α et *M. smegmatis* mc² 155. Un clone *M. smegmatis* émettant de la lumière et présentant une activité *phoA* a été sélectionné et appelé pJVED/blaF. L'insert a été retiré en utilisant *Bam*HI et la construction refermée sur elle-même, reconstruisant ainsi le pJVED_a. Afin d'obtenir le pJVED_{b,c}, le multisite de clonage a été coupé avec *Sca*I et *Kpn*I et refermé en enlevant un (pJVED_b) ou deux (pJVED_c) nucléotides du site *Sna*BI. Après fusion six cadres de lecture ont pu ainsi être obtenus. L'insert du pJVED/*hsp18* a été obtenu par amplification en chaîne par polymérase (ACP) de pPM1745 (Servant et al., 1995) en utilisant des oligonucléotides de la séquence :

- 25 18.Fw : 5'GTACCAGTACTGATCACCCGTCTCCCGCAC3'
 18.Back : AGTCAGGTACCTCGCGGAAGGGGTTCAGTGCG3'

Le produit a été coupé avec *Kpn*I et *Sca*I, et ligaturé à pJVED_a, coupé avec les mêmes enzymes, quittant ainsi le pJVED/*hsp18*.

30

Le pJVED/P19kDa et le pJVED/*erp* furent construits en coupant avec *Bam*HI l'insert de pExp410 et pExp53 respectivement, et en les insérant dans le site *Bam*HI du

multisite de clonage de pJVED_a.

Mesure de l'activité phosphatase alcaline

La présence d'activité est détectée par la couleur
bleus des colonies croissant sur un milieu de culture
contenant le substrat 5-bromo 4-chloro 3-indolyl phosphate
(XP), puis l'activité peut être mesurée quantitativement de
manière plus précise de la façon suivante :

M. smegmatis ont été cultivés dans un milieu LB
additionnés de 0,05 % de Tween 80 (Aldrich) et de
kanamycine (20 µg/ml⁻¹) à 37°C pendant 24 heures.
L'activité de la phosphatase alcaline a été mesurée par la
méthode de Brockman et Heppel (Brockman et al., 1968) dans
un extrait soniqué, avec p-nitrophénylphosphate comme
substrat de la réaction. La quantité de protéines a été
mesurée par essai Bio-Rad. L'activité phosphatase alcaline
est exprimée en unité arbitraire (densité optique à 420 nm
x µg de protéines⁻¹ x minutes⁻¹).

Mesure de l'activité luciférase

M. smegmatis a été cultivé dans un milieu LB
additionné de 0,05 % de Tween 80 (Aldrich) et de kanamycine
(20 µg/ml⁻¹) à 37°C pendant 24 heures et utilisé en pleine
croissance exponentielle (DO à 600 nm comprise entre 0,3 et
0,8). Les aliquots de suspensions bactériennes ont été
brièvement soniqués et l'extrait cellulaire a été utilisé
pour mesurer l'activité de la luciférase. 25 µl de
l'extrait soniqué ont été mélangés avec 100 µl de substrat
(système d'essai luciférase Promega) automatiquement dans
un luminomètre et la lumière émise exprimée en ULR ou RLU
(Unités Lumineuses Relatives). Les bactéries ont été
comptées par dilutions sérielles de la suspension d'origine
sur milieu agar LB kanamycine et l'activité de la
luciférase exprimée en ULR/µg de protéines bactériennes ou
en ULR/10³ bactéries.

Construction de banques génomiques de *M. tuberculosis* et
de *M. bovis*-BCG

5 Les banques ont été obtenues en utilisant
essentiellement pJVED_{a,b,c} précédemment décrits.

Préparation de macrophages issus de la moelle osseuse et
infection par *M. smegmatis* recombinants

10 Les macrophages issus de la moelle osseuse ont été
préparés comme décrits par Lang et al., 1991. En résumé,
les cellules de la moelle osseuse ont été prélevées du fémur
de souris C57BL/6 agée de 6 à 12 semaines (Iffa-Credo,
15 France). Les cellules en suspensions ont été lavées et
resuspendues dans du DMEM enrichi avec 10 % de sérum foetal
de veau, 10 % de milieu L-cell conditionné et 2 mM de
glutamine, sans antibiotiques. 10⁶ cellules ont été
ensemencées sur des plaques 24 puits Costar à fond plat
20 dans 1 ml. Après quatre jours à 37°C dans une atmosphère
humide à 10 % de teneur en CO₂, les macrophages ont été
rincés et réincubés pendant deux à quatre jours
supplémentaires. Les cellules d'un puits contrôle ont été
lysées avec du triton x 100 à 0,1 % dans l'eau et les
25 noyaux énumérés. Environ 5 x 10⁵ cellules adhérentes ont
été comptées. Pour l'infection, *M. smegmatis* portant les
différents plasmides a été cultivé en pleine phase
exponentielle (DO_{600nm} entre 0,4 et 0,8) et dilué jusqu'à
une DO de 0,1 puis 10 fois dans un milieu pour macrophage.
30 1 ml a été ajouté à chaque puits et les plaques ont été
centrifugées et incubées quatre heures à 37°C. Après trois
lavages, les cellules ont été incubées dans un milieu
contenant de l'amykacine pendant deux heures. Après trois
nouveaux lavages, les cellules infectées adhérentes ont été
35 incubées dans un milieu macrophage pendant une nuit. Les
cellules ont ensuite été lysées dans 0,5 ml de tampon de

lyse (Promega). 100 μ l ont été soniqués et la lumière émise a été mesurée sur 25 μ m. Simultanément, les bactéries ont été énumérées par étalement sur L-agar-kanamycine (20 μ g/ml⁻¹). La lumière émise est exprimée en ULR/10³ bactéries.

Analyses des banques de données

Les séquences nucléotidiques ont été comparées à EMBL et GenBank en utilisant l'algorithme FASTA et les séquences protéiques ont été analysées par similitude grâce aux banques de données PIR et Swiss Prot en utilisant l'algorithme BLAST.

Exemple 1 : Les vecteurs pJVED

Les vecteurs pJVED (Figure S1) sont des plasmides portant un gène *phoA* tronqué de *E. coli* dépourvu de codon d'initiation, de séquence signal et de séquence régulatrice. Le site multiple de clonage (SMC) permet l'insertion de fragments des gènes codants pour d'éventuelles protéines exportées ainsi que leurs séquences de régulation. Dès lors, la protéine de fusion peut être produite et présenter une activité phosphatase alcaline si elle est exportée. Seules les fusions en phase pourront être productives. Ainsi, le SMC a été modifié de sorte que les fusions peuvent être obtenues dans six phases de lecture. En aval de *phoA*, le gène *luc* de la luciférase de luciole a été inséré. Le gène complet avec le codon d'initiation mais sans qu'aucun promoteur n'ait été utilisé devrait ainsi s'exprimer avec *phoA* comme dans un opéron synthétique. Un nouveau site de liaison des ribosomes a été inséré huit nucléotides en amont du codon d'initiation de *luc*. Deux terminateurs transcriptionnels sont présents dans les vecteurs pJVED, un en amont du SMC et un second en aval de *luc*. Ces vecteurs sont des plasmides de transfert *E.*

coli-mycobacterium avec un gène de résistance à la kanamycine comme marqueur de sélection.

phoA et *luc* fonctionnent comme dans un opéron, mais
5 l'exportation est nécessaire pour l'activité *phoA*.

Quatre plasmides ont été construits par insertion dans le SMC de fragments d'ADN d'origine diverse :

Dans la première construction nommée pJVED/*blaF*, le fragment de 1,4 kb provient du plasmide déjà décrit pIA71
10 (Lim et al., 1995). Ce fragment issu du gène β -lactamase (*blaF*) de *N. fortuitum* B216 (Timm et al., 1994) inclut le promoteur muté hyperactif, le segment codant pour 32 acides aminés de la séquence signal et les 5 premiers acides aminés de la protéine mature. Ainsi cette construction
15 inclut le promoteur le plus fort connu chez *mycobacterium* et les éléments nécessaires à l'exportation de la protéine de la fusion *phoA*. Par conséquent, on peut attendre de cette construction une forte émission de lumière et une bonne activité *phoA* (cf figures 53 et 54).

20 Dans une deuxième construction nommée pJVED/*hsp18*, un fragment de 1,5 kb a été cloné à partir du plasmide déjà décrit pPM1745 (Servant et al., 1995). Ce fragment inclut les nucléotides codants pour les dix premiers acides aminés de la protéine de choc thermique de 18 kb issue de
25 *Streptomyces albus* (heat shock protein 18, HSP 18), le site de liaison du ribosome, le promoteur et, en amont, des sites régulateurs contrôlant son expression. Cette protéine appartient à la famille de alpha-crystalline de HSP à faible poids moléculaire (Verbon et al., 1992). Son
30 homologue issu de *M. leprae*, l'antigène de 18 kDa, est déjà connu pour être induit durant la phagocytose par un macrophage murin de la lignée cellulaire J-774 (Dellagostinet al., 1995). Dans des conditions de culture standard, le pJVED/*hsp18*, montre une faible activité *luc* et
35 aucune activité *phoA* (cf figures 53 et 54).

Dans une troisième construction, nommée pJVED/P19kDa, l'insert issu de pExp410 (Lim et al., 1995) a été coupé et cloné dans le SMC de pJVED_a. Ce fragment inclut les nucléotides codants pour les 134 premiers acides aminés de la protéine connue de *M. tuberculosis* 19 kDa et de ses séquences régulatrices. Comme cela a pu être mis en évidence, cette protéine est une lipoprotéine glycosylée (Carbe et al., 1993 ; Herrmann et al., 1996). Sur les figures 53 et 54, on observe, pour cette construction, une bonne activité luc correspondant à un promoteur fort, mais l'activité phoA est la plus forte des quatre constructions. L'activité phoA élevée de cette protéine de fusion avec une lipoprotéine s'explique par le fait qu'elle reste attachée à la paroi cellulaire par son extrémité N-terminal.

Dans la quatrième et dernière construction nommée pJVED/erp l'insert provient de pExp53 (Lim et al., 1995) et a été cloné dans le SMC de pJVED_a. pExp53 est le plasmide initial sélectionné pour son activité phoA et contenant une partie du gène erp de *M. tuberculosis* qui code pour un antigène de 28 kDa. Ce dernier inclut la séquence signal, une partie de la protéine mature et, en amont du codon d'initiation, le site de liaison de ribosome. Le promoteur a été cartographié. Une boîte fer (iron box) putative du type fur est présente dans cette région et encadre la région -35 du promoteur (Berthet et al., 1995). Comme prévu (figures 53 et 54) cette construction présente une bonne émission lumineuse et une bonne activité phoA. Le fait que cette protéine de fusion, contrairement à la fusion avec la lipoprotéine de 19 kDa, ne semble pas attachée à la paroi cellulaire n'exclut pas que la protéine native y soit associée. De plus, l'extrémité C-terminal de erp est absente de la protéine de fusion.

Exemple 2 : Construction d'une banque d'ADN génomique de

M. tuberculosis dans les vecteurs pJVED_g et identification d'un des membres de ces banques, (DP428), induit au cours de la phagocytose par les macrophages murins dérivés de la moelle osseuse.

5 Les différentes constructions sont testées pour leur capacité à évaluer l'expression intracellulaire des gènes identifiés par l'expression de *phoA*. Dans cet objectif, l'activité luc est exprimée en URL pour 10³ bactéries en culture axénique et/ou dans des conditions
10 intracellulaires. L'induction ou la répression suivant la phagocytose par les macrophages murins dérivés de la moelle osseuse peut être évaluée convenablement par la mesure des activités spécifiques. Les résultats de deux expériences distinctes sont présentés dans le tableau 2.

15 Le plasmide pJVED/*hsp18* a été utilisé comme contrôle positif pour l'induction durant la phase de croissance intracellulaire. Bien que l'induction du promoteur par le chauffage de la bactérie à 42°C n'ait pas été concluant la phagocytose de la bactérie conduit
20 clairement à une augmentation de l'activité du promoteur. Dans toutes les expériences, l'activité luc intracellulaire a été fortement induite, augmentant de 20 à 100 fois l'activité basale initialement faible (Servant, 1995).

Le plasmide pJVED/*blaF* a été utilisé comme contrôle
25 de la modulation non spécifique au cours de la phagocytose. De faibles variations ont pu être mises en évidence, probablement dues à des changements de conditions de cultures. Quoi qu'il en soit, ces faibles variations ne sont pas comparables à l'induction observée avec le
30 plasmide pJVED/*hsp18*.

Tous les membres de la banque d'ADN ont été testés par mesure de l'activité du promoteur durant la croissance intracellulaire. Parmi eux, le DP428 est fortement induit au cours de la phagocytose (tableaux 1 et 2).

TABLEAU 1

Construction	% Récupération	URL/10 ³ bactéries extracellulaire	URL/10 ³ bactéries intracellulaire	Induction
pJVED/ <i>hla</i> +	0,5	1460	1727	1,2
pJVED/ <i>hsp18</i>	0,6	8	57	7,1
pJVED/DP428	0,7	0,06	18	100
Construction	% Récupération C57BL/6 Balb/C	URL/10 ³ bactéries extracellulaire	URL/10 ³ bactéries intracellulaire C57BL/6 Balb/C	Induction C57BL/6 Balb/C
pJVED/ <i>hla</i> +	7 1,1	662	250 911	0,4 1,4
pJVED/ <i>hsp18</i>	6,7 1,7	164	261 325	1,6 2
pJVED/DP428	1,6 2,1	0,08	1,25 3,3	15,6 41

5 TABLEAU 2

Construction	% Récupération	URL/10 ³ bactéries extracellulaire	URL/10 ³ bactéries intracellulaire	Induction
pJVED/ <i>hla</i> +	22	1477	367	0,25
pJVED/ <i>hsp18</i>	7	0,26	6,8	26
pJVED/DP428	21	0,14	4	28

Le fragment nucléotidique codant pour la région N-terminale du polypeptide DP428 de séquence SEQ ID N° 28 est contenu dans le plasmide déposé à la CNCM sous le N° I-1818.

La totalité de la séquence codant pour le polypeptide DP428 a été obtenue comme détaillée ci-après.

Une sonde a été obtenue par PCR à l'aide des oligonucléotides de séquence SEQ ID N° 25 et SEQ ID N° 26. Cette sonde a été marquée par extension aléatoire en présence de ³²P dCTP. Une hybridation de l'ADN génomique de *M. tuberculosis* souche Mt103 préalablement digéré par l'endonucléase ScaI a été réalisée à l'aide de ladite

sonde. Les résultats de l'hybridation ont fait apparaître qu'un fragment d'ADN d'environ 1,7 kb était marqué. Du fait qu'il existe un site *ScaI* s'étendant du nucléotide nt 984 au nucléotide nt 989 de la séquence SEQ ID N° 1, c'est-à-dire du côté 5' de la séquence utilisée comme sonde, la fin de la séquence codante est nécessairement présente dans le fragment détecté par hybridation.

L'ADN génomique de la souche Mt 103 de *M. tuberculosis*, après digestion par *ScaI*, a subi une migration sur un gel d'agarose. Les fragments de tailles comprises entre 1,6 et 1,8 kb ont été clonés dans le vecteur pSL1180 (Pharmacia) préalablement clivé par *ScaI* et déphosphorylé. Après transformation de *E. coli* avec les vecteurs recombinants résultants, les colonies obtenues ont été criblées à l'aide de la sonde. Le criblage a permis d'isoler six colonies hybridant avec cette sonde.

Les inserts contenus dans les plasmides des clones recombinants précédemment sélectionnés ont été séquencés, puis les séquences alignées de manière à déterminer la totalité de la séquence codant pour DP428, plus spécifiquement la SEQ ID N° 2.

Un couple d'amorces a été synthétisé afin d'amplifier, à partir de l'ADN génomique de *M. tuberculosis*, souche Mt 103, la totalité de la séquence codant pour le polypeptide DP428. L'amplicon obtenu a été cloné dans un vecteur d'expression.

Des couples d'amorces appropriés pour l'amplification et le clonage de la séquence codant pour le polypeptide DP428 peuvent être aisément réalisés par l'homme du métier, sur la base des séquences nucléotidiques SEQ ID N°1 et SEQ ID N°2.

Un couple d'amorces particulier selon l'invention est le couple d'amorces suivantes, capable d'amplifier l'ADN codant pour le polypeptide DP428 dépourvu de sa séquence signal ;

- Amorce aller (SEQ ID N° 29), comprenant la séquence allant du nucléotide en position nt 1021 au nucléotide nt 1044 de la séquence SEQ ID N° 2 :

5' -AGTGCATGCTGCTGGCCGAACCATCAGCGAC- 3'

- Amorce retour (SEQ ID N° 30), comprenant la séquence complémentaire de la séquence allant du nucléotide en position nt 1345 au nucléotide en position nt 1325 de la séquence SEQ ID N° 2 :

5' -CAGCCAGATCTGCGGGCGCCTGACACCGCCTG- 3',

dans lesquelles la partie soulignée représente les séquences hybridant spécifiquement avec la séquence SEQ ID N° 2 et les extrémités 5' correspondent à des sites de restriction en vue du clonage de l'amplicon résultant dans un vecteur de clonage et/ou d'expression.

Un vecteur particulier utilisé pour l'expression du polypeptide DP428 est le vecteur pQE70 commercialisé par la société Qiagen.

Exemple 3 : La séquence complète du gène DP428 et de ses régions flanquantes.

Une sonde de la région codante de DP428 a été obtenue par ACP, et utilisée pour hybrider l'ADN génomique de différentes espèces de mycobactéries. D'après les résultats de la figure 3, le gène est présent uniquement dans les mycobactéries du complexe de *M. tuberculosis*.

L'analyse de la séquence suggère que DP428 pourrait faire partie d'un opéron. La séquence codante et les régions flanquantes ne présentent aucune homologie avec des séquences connues déposées dans les banques de données.

D'après la séquence codante, ce gène code pour une protéine de 10 kDa avec un peptide signal, une extrémité C-terminal hydrophobe terminée par deux arginines et précédée par un motif LPISG semblable au motif connu LPXTG. Ces deux

arginines pourraient correspondre à un signal de rétention et la protéine DP428 pourrait être accrochée par ce motif à des peptidoglycanes comme cela a déjà été décrit chez d'autres bactéries Gram⁺ (Navarre et al., 1994 et 1996).

5 Le mécanisme de survie et de croissance intracellulaire des mycobactéries est complexe et les relations intimes entre la bactérie et la cellule hôte restent inexpliquées. Quel que soit le mécanisme, la croissance et la survie intracellulaire des mycobactéries
10 dépend de facteurs produits par la bactérie et capables de moduler la réponse de l'hôte. Ces facteurs peuvent être des molécules exposées à la surface cellulaire telle que LAM ou des protéines associées à la surface cellulaire, ou des molécules activement secrétées.

15 D'un autre côté, intracellulairement, les bactéries elles-mêmes doivent faire face à un environnement hostile. Elles semblent y répondre par des moyens proches de ceux mis en oeuvre dans les conditions de stress, par l'induction de protéines de choc thermique (Dellagostin et
20 al., 1995), mais aussi par induction ou la répression de différentes protéines (Lee et al., 1995). En utilisant une méthodologie dérivée de la PCR, Plum et Clark-curtiss (Plum et al., 1994) ont montré qu'un gène de *M. avium* inclut dans un fragment d'ADN de 3 kb, est induit après la phagocytose
25 par des macrophages humains. Ce gène code pour une protéine exportée comprenant une séquence leader mais ne présentant pas d'homologie significative avec les séquences proposées par les banques de données. L'induction, pendant la phase de croissance intracellulaire, d'une protéine de choc
30 thermique de faible poids moléculaire issue de *M. leprae* a également été mise en évidence (Dellagostin et al., 1995). Dans une autre étude, les protéines bactériennes de *M. tuberculosis* ont été métaboliquement marquées pendant la phase de croissance intracellulaire ou bien dans des
35 conditions de stress et séparées par électrophorèse sur gel à deux dimensions : 16 protéines de *M. tuberculosis* ont été

induites et 28 réprimées. Les mêmes protéines sont mises en jeu au cours de stress provoqué par un faible pH, un choc thermique, H₂O₂, ou au cours de la phagocytose par des monocytes humains de la lignée THP1. Quoi qu'il en soit, le comportement des protéines induites et réprimées était unique dans chaque condition (Lee et al., 1995). Pris ensemble, ces résultats indiquent qu'un dialogue moléculaire subtile est mis en place entre les bactéries et leurs hôtes cellulaires. De ce dialogue dépend probablement le sort de l'organisme intracellulaire.

Dans ce contexte, l'induction de l'expression de DP428 pourrait être d'une importance majeure, indiquant un rôle important de cette protéine dans la survie et la croissance intracellulaire.

La méthode utilisée dans ces expériences pour évaluer l'expression intracellulaire des gènes (cf. Jacobs et al., 1993, pour la méthode de détermination de l'expression de la luciférase de luciole, et Lim et al., 1995, pour la méthode de détermination de l'expression du gène *PhoA*) présente l'avantage d'être simple comparée aux autres techniques comme la technique décrite par Mahan et al. (Mahan et al., 1993) adaptée aux mycobactéries et proposée par Bange et al. (Bange et al., 1996), ou la méthode subtractive basée sur l'ACP décrite par Plum et Clark-curtiss (Plum et al., 1994). Il existe indiscutablement une variabilité comme le montre la comparaison des différentes expériences. Bien que provoquer l'induction ou la répression soit suffisant, il est désormais possible de l'évaluer fournissant ainsi un outil supplémentaire d'études physiologiques des protéines exportées identifiées par fusion avec *phoA*.

Exemple 4 :

Recherche d'une modulation de l'activité des promoteurs lors des phases intramacrophagiques.

Des macrophages de moelle osseuse de souris sont préparés comme décrit par Lang et Antoine (Lang et al., 1991). Les bactéries de *M. segmentis* recombinantes, dont on a déterminé l'activité luciférase par 10^3 bactéries comme précédemment, sont incubées à 37°C sous atmosphère humidifiée et enrichie en CO₂ à 5%, pendant 4 heures en présence de ces macrophages de telle manière qu'elles soient phagocytées. Après rinçage pour éliminer les bactéries extracellulaires restantes, on ajoute au milieu de culture de l'amikacine (100 µg/ml) pendant deux heures. Après un nouveau rinçage, le milieu est remplacé par un milieu de culture (DMEM enrichi de 10 % de sérum de veau et 2 mM de glutamine) sans antibiotiques. Après une nuit d'incubation comme précédemment, les macrophages sont lysés à froid (4°C) à l'aide d'un tampon de lyse (cse lysis buffer, Promega), et l'activité luciférase par 10^3 bactéries déterminée. Le rapport des activités à la mise en culture et après une nuit donne le coefficient d'induction.

Exemple 5 :

Isolément d'une série de séquences par séquençage directement à partir des colonies.

Une série de séquences permettant l'expression et l'exportation de *phoA* ont été isolées à partir de l'ADN de *M. Tuberculosis* ou de *M. Bovis* BCG. Parmi ce groupe de séquences, deux d'entre elles ont été d'avantage étudiées, les gènes entiers correspondant aux inserts ont été clonés, séquencés, et des anticorps contre le produit de ces gènes ont servi à montrer en microscopie électronique que ces gènes codaient pour des antigènes retrouvés à la surface des bacilles de la tuberculose. L'un de ces gènes encodant pour une séquence signal d'exportation consensus, l'autre des ne possédait aucune caractéristique de gène codant pour une protéine exportée, d'après la séquence. Un

autre gène DP428 a été séquencé avant que la séquence du génome de *M. Tuberculosis* ne soit disponible. Il contient une séquence ressemblant à la séquence consensus d'attachement au peptidoglycane, ce qui suggère qu'il s'agit aussi d'un antigène vraisemblablement retrouvé à la surface des bacilles de la tuberculose. L'étude des trois gènes *erp*, *des*, et celui codant pour DP428 montre que le système *phoA* que nous avons développé chez les mycobactéries permet de repérer des gènes codant pour des protéine exportées sans déterminant repérable par des études *in silico*. Ceci est particulièrement vrai pour les polypeptides qui ne possèdent pas de séquence signal consensus (*des*) ou non pas de similarité avec des protéines de fonction connue (*erp* et DP428).

Un certain nombre d'inserts ont été identifiés et séquencés avant la connaissance du génome de *M. Tuberculosis*, d'autres après. Ces séquences peuvent être considérées comme des amorces permettant de rechercher des gènes codant pour des protéines exportées. A ce jour, une série d'amorces ont été séquencées et les gènes entiers correspondants ont été soit séquencés, soit identifiés d'après la séquence publiée du génome. Pour tenir compte des erreurs de séquençage toujours possibles, les régions en amont ou en aval de certaines amorces ont été considérées comme pouvant faire partie de séquences codant pour des protéines exportées. Dans certains cas des similarités avec des gènes codant pour des protéines exportées ou des séquences caractéristiques de signaux d'exportation ou des caractéristiques topologiques de protéines membranaires ont été détectées.

Des séquences amorces s'avèrent correspondre à des gènes appartenant à des familles de gènes possédant plus de 50 % de similarité. On peut ainsi indiquer que les autres gènes détectés par similarité avec une amorce codent pour

des protéines exportées. C'est le cas de la séquence SEQ ID N° 8G et SEQ ID N° 8H possédant plus de 77 % de similarité avec SEQ ID N° 8A'.

Les séquences pouvant coder pour des protéines exportées sont les suivantes : SEQ ID N° 1, 8, 9, 8G, 8H, 13, 3, 10, 19, 20, 6, 16, 22, 23, 24, 39, 44, 46, et 50.

Des gènes identifiés d'après les amorces à partir de la séquence du génome n'ont aucune caractéristique (d'après la séquence) de protéines exportées. Il s'agit des séquences suivantes : SEQ ID N° 4, 27, 11, 12, 14, 7, 15, 17, 18, 21, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 45, 47, 48, et 49.

D'après la séquence d'autres organismes comme *E. coli*, on peut rechercher dans la séquence du génome de *M. tuberculosis*, des gènes possédant des similarités avec des protéines connues pour être exportées chez d'autres organismes bien que ne possédant pas de séquence signal d'exportation. Dans ce cas une fusion avec *phoA* est un protocole avantageux pour déterminer si ces séquences de *M. tuberculosis* codent pour des protéines exportées bien que ne présentant pas de séquence signal consensus. Il a été en effet possible de cloner SEQ ID N° 49, une séquence similaire à un gène de *E. coli* de la famille *htrA*. Une fusion de SEQ ID N° 49 avec *phoA* conduit à l'expression et à l'exportation de *phoA*. Des colonies *M. smegmatis* hébergeant une fusion SEQ ID N° 49 *phoA* sur un plasmide pJVED sont bleues.

SEQ ID N° 49 est donc considérée comme une protéine exportée.

La méthode *phoA* est donc utile pour détecter d'après la séquence de *M. Tuberculosis* des gènes codant pour des protéines exportées sans qu'ils ne codent pour des séquences caractéristiques des protéines exportées.

Même si une séquence possède des déterminants de protéines exportées, cela ne démontre pas une exportation fonctionnelle. Le système *phoA* permet de montrer que le gène suspecté code réellement pour une protéine exportée.

- 5 Ainsi, il a été vérifié que la séquence SEQ ID N° 50 possédait bien des signaux d'exportation.

TABLEAU 3

SEQ ID N°	Référence de la séquence correspondante prédite par Cole et al.		Annotation
SEQ ID N°1	Rv 0203	*	Séquence hydrophobe en N-terminal
SEQ ID N°4			
SEQ ID N°27	Rv 2050		Pas de prédiction
SEQ ID N°8			
SEQ ID N°9	Rv 2563	*	Protéine membranaire
SEQ ID N°			
SG',H'	Rv 0072	*	Possible protéine de transport transmembranaire de type ABC
SEQ ID N°11	Rv 0546c	ML	Protein S-D Lactoyl Glutathione-méthyl glyoxal lyase
SEQ ID N°12	pas de prédiction		non retrouvé dans <i>M.tuberculosis</i> H37rv
SEQ ID N°13			
SEQ ID N°3	Rv 1984c	*	probable précurseur cutinase avec une séquence signal N-terminale
SEQ ID N°14			
SEQ ID N°7	pas de prédiction		pas de prédiction
SEQ ID N°15	avec décalage de lecture, pourrait être en phase avec Rv 2530c		pas de prédiction
SEQ ID N°17	Rv 1303	ML	pas de prédiction
SEQ ID N°18	Rv 0199	ML	pas de prédiction
SEQ ID N°19	Rv 0418	*	site de fixation de lipoprotéine membranaire procaryote, similarité avec la N-acétyl putromycine acétyl hydrolase
SEQ ID N°20	Rv 3576	*	contient un site de fixation de lipoprotéine membranaire procaryote, similarité avec une
SEQ ID N°6			

			sérine/thréonine protéine kinase
SEQ ID N° 21	Rv 3365c	ML	similarité avec une métallo peptidase à zinc
SEQ ID N° 31	non prédite		pas de prédiction
SEQ ID N° 32	Rv 0822c	ML	Existence d'une région consensus avec la famille drac
SEQ ID N° 33	Rv 1044		pas de prédiction
SEQ ID N° 34	non prédite		pas de prédiction
SEQ ID N° 35	Rv 2169c		pas de prédiction
SEQ ID N° 36	Rv 3909	ML	pas de prédiction
SEQ ID N° 37	Rv 2753c		similarité avec des dihydropricolinate synthases
SEQ ID N° 38	Rv 0175		pas de prédiction
SEQ ID N° 39	Rv 3006	ML	prédiction de séquence signal de lipoprotéine
SEQ ID N° 40	Rv 0549c		pas de prédiction
SEQ ID N° 41	Rv 2975c pouvant être en phase avec Rv 2974c		similarité avec protéine de subtilis
SEQ ID N° 42	Rv 2622		similarité avec une méthyl transférase
SEQ ID N° 43	Rv 3278c	ML	pas de prédiction
SEQ ID N° 44	Rv 0309	*	pas de prédiction
SEQ ID N° 45	Rv 2169c	ML	pas de prédiction
SEQ ID N° 46	Rv 1411c	*	probable lipoprotéine avec une séquence signal N-terminale
SEQ ID N° 47	Rv 1714		similarité avec une gluconate 3-déhydrogénase
SEQ ID N° 48	Rv 0331		similarité avec une sulfide déhydrogénase et une sulfide quinone réductase
SEQ ID N° 49	Rv 0983	ML	Similarité avec une sérine protéase HtrA

SEQ ID N°5			
SEQ ID N°16	Rv 3810	*	Protéine de surface Bertholet et al. 1995
SEQ ID N°22 SEQ ID N°23 SEQ ID N°24	Rv 3763	*	Contient un site de fixation de lipoprotéine membranaire eucaryote
SEQ ID N°50	Rv 0125	*	Site actif des sérines protéases Séquence signal N-terminale possible

Légende du tableau 3 :

Correspondance des séquences selon l'invention avec les
séquences prédites par Cole et al. 1998, Nature, 393, 537-
544.

* : Prédiction que la protéine codée par la séquence
soit exportée

ML : Prédiction de similarité avec *M. leprae*.

Exemple 6 :

Caractéristiques et obtention de la protéine M1C25

L'extrémité N terminale de la protéine M1C25 a été
détectée par le système PhoA comme permettant l'exportation
de la protéine de fusion, nécessaire à l'obtention de son
activité phosphatase.

La séquence d'ADN codant pour l'extrémité N terminale de
la protéine M1C25 est contenue dans la séquence SEQ ID N°
20 de la présente demande de brevet.

A partir de cette séquence amorce, le gène complet
codant pour la protéine M1C25 a été recherché dans le
génome de *M. tuberculosis* (Fondation Wellcome Trust, site
Sanger).

Le centre Sanger a attribué à M1C25 les noms:

Rv3576,
MTCY06G11.23,
pknM

5 Séquence SEQ ID N° 29 du gène complet M1C25 (714 bases):
cf. Figure 29

Ce gène code pour une protéine de 237 AA, de 25 kDa de
masse molaire. Cette protéine est référencée dans les
10 banques sous les appellations:
PID:e306716,
SPTREMBL:P96858

Séquence SEQ ID N° 30 de la protéine M1C25 (237 acides
15 aminés): cf. Figure 30

M1C25 contient un site de fixation à la partie
lipidique des lipoprotéines de membranes des procaryotes
(PS00013 Prokaryotic membrane lipoprotein lipid attachment
20 site:
CTGGTCGGTG CGTGCATGCT CGCAGCCCGGA TCC).

La fonction de M1C25 n'est pas certaine mais elle
possède très probablement une activité "sérine/thréonine-
protéine kinase". Des ressemblances sont à noter avec la
25 moitié C terminale de K08G_MYCTU Q11051 Rv1266c
(MTCY50.16). Des similarités sont aussi retrouvées avec
KY28_MYCTU.

En 5' du gène codant pour M1C25 se trouve un gène
codant potentiellement pour une protéine régulatrice
30 (PID:e306715, SPTREMBL:P96857, Rv3575c, (MTCY06G11.22c))

Le profil d'hydrophobicité (Kyte et Doolittle) de M1C25
est représenté à la figure 56.

Un site de clivage de la séquence signal est prédit
(SignalP V1.1; World Wide Web Prediction Server, Center for
35 Biological Sequence Analysis) entre les acides aminés 31 et
32: AVA-AD. Ce site de coupure est derrière un motif "AXA"
classique. Cette prédiction est compatible avec le profil

d'hydrophobicité. Dans cette séquence signal potentielle il est a remarqué la répétition trois fois de la séquence des trois acides aminés LAA.

- 5 Clonage du gène M1C25 en vue de la production de la protéine qu'il code:

Un couple d'amorces a été synthétisé afin d'amplifier, à partir de l'ADN génomique de *M. tuberculosis*, souche
10 H37Rv, la totalité de la séquence codant pour le polypeptide M1C25. L'amplicon obtenu a été cloné dans un vecteur d'expression.

Des couples d'amorces appropriés pour l'amplification
15 et le clonage de la séquence codant pour M1C25 ont été synthétisés :

-amorce aller :

5' -ATAATACCATGGGCAAGCAGCTAGCCGCGC- 3'

-amorce retour :

20 5' -ATTTATAGATCTCTGCTTAGCAACCTTGSCCGCG- 3'

La partie soulignée représente les séquences hybridant spécifiquement avec la séquence M1C25 et les extrémités 5' correspondant à des sites de restriction en vue du clonage
25 de l'amplicon résultant dans un vecteur de clonage et/ou d'expression.

Un vecteur particulier utilisé pour l'expression du polypeptide M1C25 est le vecteur pQE60 commercialisé par la société Qiagen, en suivant le protocole et les
30 recommandations proposées par cette marque.

Les cellules utilisées pour le clonage sont des bactéries : *E. coli* XL1-Blue (résistante à la tétracycline).

Les cellules utilisées pour l'expression sont des
35 bactéries : *E. coli* M15 (résistante à la kanamycine) contenant le plasmide pRep4 (M15 pRep4).

La production de la protéine MC25 est illustrée par les figures 57 A et B. (Extraits bactériens de la souche E. coli M15 contenant le plasmide pM1C25. Les cultures bactériennes et les extraits sont préparés selon Sambrook et al. (1989). L'analyse des extraits bactériens est effectuée selon les instructions de Quiagen (1997).

Références bibliographiques

- AIDS therapies, 1993, in *Mycobacterial infections*, ISBN 0-9631698-1-5, pp. 1-11.
- 5 Altschul, S.F. et al., 1990, *J. Mol. Biol.*, 215 : 403-410.
- Andersen, P. et al., 1991, *Infect. Immun.*, 59 :1905-1910.
- Andersen, P. et al., 1995, *J. Immunol.*, 154, 3359-3372.
- Bange, F.C., A.M. Brown, and W.R. Jacobs JR., 1996, Leucine auxotrophy restricts growth of *Mycobacterium bovis* BCG in
- 10 macrophages. *Infect. Immun.*, 64, : 1794-1799.
- Barany, F., 1911, *Proc. Natl. Acad. Sci. USA*, 88 :189-193.
- Bates, J. et al., 1986, *Am. Rev. Respir. Dis.*, 134 :415-417.
- Bates, J. 1979. *Chest*, 76(Suppl.):757-763.
- 15 Bates, J. et al., 1986. *Am. Rev. Respir. Dis.* 134 :415-417.
- Berthet, F.X., J. Rauzier, E.M. Lim, W. Philipp, B. Gicquel, and D. Portnoi, 1995, Characterization of the *M. tuberculosis* *erp* gene encoding a potential cell surface protein with repetitive structures. *Microbiology*. In press
- 20 Borremans, M. et al., 1989, *Biochemistry*, 7 : 3123-3130.
- Bouvet, E. 1994. *Rev. Fr. Lab.* 273 :53-56.
- Brockman, R.W. and Heppel L.A., 1968, On the localization of alkaline phosphatase and cyclic phosphodiesterase in *Escherichia coli*, *Biochemistry*, 7 : 2554-2561.
- 25 Burg, J.L. et al., 1996, *Mol. and Cell. Probes*, 10 :257-271.
- Chevrier, D. et al., 1993, *Mol. and Cell. Probes*, 7 :187-197.
- Clemens, D.L., 1996, Characterization of the *Mycobacterium*
- 30 *tuberculosis* phagosome, *Trends Microbiol.*, 4 : 113-118.
- Chu ,B.C.F. et al., 1986, *Nucleic Acids Res.*, 14 :5591-5603.

- Clemens, D.L. and Horwitz M.A., 1995, Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited, *J. Exp. Med.*, 181 : 257-270.
- 5 Colignon J.E., 1996. Immunologic studies in humans. Measurement of proliferative responses of cultured lymphocytes. *Current Protocols in Immunology*, NIH, 2, Section II.
- Daniel, T.M. et al. 1987. *Am. Rev. Respir. Dis.*, 135 :1137-1151).
- 10 Dellagostin, O.A., Esposito G., Eales L.-J., Dale J.W. and McFadden J.J., 1995, Activity of mycobacterial promoters during intracellular and extracellular growth. *Microbiol.*, 141 : 2123-2130.
- 15 Drake, T.A. et al. 1987. *J. Clin. Microbiol.* 25:1442-1445.
- Dramsfi et al., 1997, *Infection and Immunity*, 65, 5 : 1615-1625.
- Duck, P. et al., 1990, *Biotechniques*, 9:142-147.
- Erlich, H.A. 1989. In *PCR Technology. Principles and Applications for DNA Amplification*. New York: Stockton Press.
- 20 Felgner et al., 1987, *Proc. Natl. Acad. Sci.*, 84:7413.
- Fraley et al., 1980, *J. Biol. Chem.*, 255:10431.
- Gaillard, J.L., Berche P., Frehel C., Gouin E. and Cossart P., 1991, Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from Gram-positive cocci, *Cell.*, 65 : 1127-1141.
- 25 Garbe, T., Harris D., Vordermeir M., Lathigra R., Ivanyi J. and Young D., 1993, Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation, *Infect. Immun.*, 61 : 260-267.
- 30

- Guatelli, J.C. et al., 1990, Proc. Natl. Acad. Sci. USA, 87:1874-1878.
- Harboe et al., 1996, Infect. Immun., 64, 16-22.
- Herrmann, J.L., O'Gaora P., Gallagher A., Thole J.E.R. and
5 Young D.B., 1996, Bacterial glycoproteins: a link between
glycosylation and proteolytic cleavage of a 19 kDa antigen
from *Mycobacterium tuberculosis*, EMBO J. 15 : 3547-3554.
- Houbenweyl, 1974, in *Methoden der Organischen Chemie*, E.
Wunsch Ed., Volume 15-I et 15-II, Thieme, Stuttgart.
- 10 Huygen, K. et al., 1996, Nature Medicine, 2(8):893-898.
- Innis, M.A. et al., 1990, in *PCR Protocols. A guide to
Methods and Applications*. San Diego: Academic Press.
- Isberg, R.R., Voorhis D.L. and Falkow S., 1987,
Identification of invasins: a protein that allows enteric
15 bacteria to penetrate cultured mammalian cells, Cell, 50 :
769-778.
- Jacobs, W.R. et al., 1991. Construction of mycobacterial
genomic libraries in shuttle cosmids. Genetic Systems for
Mycobacteria, Methods in Enzymology, 204 : 537-555.
- 20 Jacobs, W.R. et al., 1993, Science, 260 :819-822.
- Kaneda, et al., 1989, Science, 243:375
- Kiehn, T.E., et al. 1987. J. Clin. Microbiol. 25 :1551-
1552.
- Kievitis ,T. et al., 1991, J. Virol. Methods, 35 :273-286.
- 25 Kohler, G. et al., 1975, Nature, 256(5517):495-497.
- Kwoh, D.Y. et al., 1989, Proc. Natl. Acad. Sci. USA, 86
:1173-1177.
- Landegren ,U. et al., 1988, Science, 241,:1077-1080.
- Lang, T. and Antoine J.-C., 1991, Localization of MHC
30 classII molecules in murine bone marrow-derived
macrophages. Immunology, 72 : 199-205.

- Lee, B.Y. and Horwitz M.A., 1995, Identification of macrophage and stress-induced proteins of *Mycobacterium tuberculosis*, J. Clin. Invest., 96 : 245-249.
- Lim, E.M., Rauzier J., Timm J., Torrea G., Murray A.,
5 Gicquel B. and Portnoi D., 1995, Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins, using *phoA* gene fusions, J. Bacteriol., 177 : 59-65.
- Lizardi, P.M. et al., 1988, Bio/technology, 6 :1197-1202.
- 10 Mahan, M.J. et al., 1993, Selection of bacterial virulence genes that are specifically induced in host tissues, Science, 259 : 686-688.
- Manoil L., Mekolans J.J. and Beckwith J., J. Bacteriol., 1990, 172, 515-518.
- 15 Matthews, J.A. et al., 1988, Anal. Biochem., 169:1-25.
- Merrifield, R.D., 1966, J. Am. Chem. Soc., 88(21):5051-5052.
- Midoux, 1993, Nucleic Acids Research, 21:871-878/
- Miele, E.A. et al., 1983, J. Mol. Biol., 171:281-295.
- 20 Minton, N.P., 1984, Gene, 31, 269-273.
- Montgomery et al., 1993, DNA Cell Biol., 12:777-783.
- Navarre, W.W. et al., 1994, Molecular Microbiologie, 14(1):115-121.
- Navarre, W.W. et al., 1996, J. of Bacteriology, 178, 2 :441-
25 446.
- Pagano et al., 1967, J. Virol., 1 :891
- Pastore, 1994, Circulation, 90:I-517.
- Patel, et al. 1990. J. Clin. Microbiol. :513-518.
- Prentki, B. et Krish H.M., 1984, Gene, 29 : 303-313.
- 30 Pettersson R., Nordfelth J., Dubinina E., Bergman T., Gustafsson M., Magnusson K.E. and Wolf-Watz H., 1996,

- Modulation of virulence factor expression by pathogen target cell contact. *Science.*, 273 : 1231-1233.
- Plum, G. and Clark-Curtiss J.E., 1994, Induction of *Mycobacterium avium* gene expression following phagocytosis
- 5 by human macrophages. *Infect. Immun.*, 62 : 476-483.
- Roberts, M.C., et al. 1987. *J. Clin. Microbiol.* 25, :1239-1243.
- Rolfs, A. et al. 1991. In *PCR Topics. Usage of Polymerase Chain reaction in Genetic and Infectious Disease*. Berlin:
- 10 Springer-Verlag.
- Sambrook, J. et al. 1989. In *Molecular cloning : A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sanchez-Pescador, R., 1988, *J. Clin. Microbiol.*,
- 15 26(10), :1934-1938.
- Schneewind, O. et al., 1995, *Science*, 268 : 103-106.
- Segev D., 1992, in « Non-radioactive Labeling and Detection of Biomolecules », Kessler C. Springer Verlag, Berlin, New-York, 197-205.
- 20 Servant, P. and Mazodier P., 1995, Characterization of *Streptomyces albus* 18-kilodalton heat shock-responsive protein. *J. Bacteriol.*, 177 : 2998-3003.
- Shiver, J.W., 1995, in *Vaccines 1995*, eds Chanock, R.M. Brown, F. Ginsberg, H.S. & Norrby, E.), pp.95-98, Cold
- 25 Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sorensen et al., 1995, *Infect. Immun.*, 63, 1710-1717.
- Stone, B.B. et al., 1996, *Mol. and Cell. Probes*, 10 :359-370.
- 30 Stover, C.K., Bansal G.P., Hansen M.S., Burlein S.R., Palaszynski S.R., Young J.F., Koenig S., Young D.B., Sadziene A. and Barbour A.G., 1993, Protective immunity elicited by recombinant Bacille Calmette-Guérin (BCG)

- expressing outer surface protein A (OspA) lipoprotein: a candidate Lyme disease vaccine. *J. Exp. Med.*, 178 : 197-209.
- Sturgill-Koszycki, S., Schlesinger P.H., Chakroborty P.,
5 Haddix P.L., Collins H.L., Fok A.K., Allen R.D., Gluck S.L., Heuser J. and Russell D.G., 1994, Lack of acidification in *Mycobacterium* phagosomes by exclusion of the vesicular proton-ATPase. *Science.*, 263 : 678-681.
- Tascon, R.E et al., 1996, *Nature Medicine*, 2(8):888-892.
- 10 Technique assemblage oligonucléotides, 1983, *Proc. Natl. Acad. Sci. USA*, 80 : 7461-7465.
- Technique des bêta-cyanéthylphosphoramidites, 1986, *Bioorganic Chem.*, 4 : 274-325.
- Thierry, D. et al., 1990, *Nucl. Acid Res.*, 18 : 188.
- 15 Timm, J., Perilli M.G., Duez C., Trias J., Orefici G., Fattorini L., Amicosante G., Oratore A., Boris B., Frere J.M., Pugsley A.P. and Gicquel B., 1994, Transcription and expression analysis, using *lacZ* and *phoA* gene fusions, of *Mycobacterium fortuitum* B-lactamase genes cloned from a
20 natural isolate and a high-level B-lactamase producer. *Mol. Microbiol.*, 12 : 491-504.
- Tuberculosis Prevention Trial, 1980, Mendis, « Trial of BCG vaccines in South India for Tuberculosis Infection », *Indian Journal of Medical research*, 1972 (Suppl.):1-74.
- 25 Urdea, M.S. et al., 1991, *Nucleic Acids Symp. Ser.*, 24 : 197-200.
- Urdea, M.S., 1988, *Nucleic Acids Research*, 11: 4937-4957.
- Verbon, A., Hartskeerl R.A., Schuitema A., Kolk A.H., Young D.B. and Lathigra R., 1992, The 14,000-molecular-weight
30 antigen of *Mycobacterium tuberculosis* is related to the alpha-crystallin family of low-molecular-weight heat shock proteins. *J Bacteriol.*, 174 : 1352-1359.

- Walker, G.T. et al., 1992, Nucleic Acids Res., 20:1691-1696.
- Walker, G.T. et al., 1992, Proc. Natl. Acad. Sci. USA, 89:392-396.
- 5 Wiker, H.G. et al., 1992, Microbiol. Rev., 56 :648-661.
- Yamaguchi, R. et al., 1989, Infect. Immun., 57 :283-288 ;
- Xu, S., Cooper A., Sturgill-Koszycki S., van Heyningen T., Chatterjee D., Orme I., Allen P. and Russel D.G., 1994, Intracellular trafficking in Mycobacterium tuberculosis and
- 10 Mycobacterium avium-infected macrophages, J. Immuno., 153: 2568-2578.
- Young, D.B. et al., 1992, Mol. Microbiol., 6 :133-145.
- Yuen, L.K.W. et al., 1993, J. Clin. Microbiol., 31 : 1615-1618.

REVENDICATIONS

1. Vecteur recombinant de criblage, de clonage et/ou d'expression, caractérisé en ce qu'il se réplique chez des mycobactéries et en ce qu'il contient ;
- 5 1) un réplicon fonctionnel chez les mycobactéries ;
2) un marqueur de sélection ;
3) une cassette rapporteur comprenant ;
a) un site de clonage multiple (polylinker),
10 b) éventuellement un terminateur de transcription actif chez les mycobactéries, en amont du polylinker,
c) une séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression, d'exportation et/ou de sécrétion de protéine, ladite séquence nucléotidique étant dépourvue de son codon d'initiation et de ses
15 séquences de régulation, et
d) une séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de promoteurs contenus dans le même fragment, ladite séquence nucléotidique étant
20 pourvue de son codon d'initiation.
2. Vecteur recombinant selon la revendication 1, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression,
25 d'exportation et/ou de sécrétion de protéine est une séquence codante issue du gène *phoA* de la phosphatase alcaline.
3. Vecteur recombinant selon l'une des revendications 1 et 2, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression,
30 d'exportation et/ou de sécrétion de protéine est une séquence codante du gène de la β -agarase, de la nucléase d'un staphylocoque ou de la β -lactamase d'une mycobactérie.

4. Vecteur recombinant selon l'une des revendications 1 à 3, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de promoteurs contenus dans le même fragment est une séquence
5 codante issue du gène *luc* de la luciférase de luciole.
5. Vecteur recombinant selon l'une des revendications 1 à 4, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de
10 promoteurs contenus dans le même fragment est une séquence codante issue du gène *GFP* de la Green Fluorescent Protein.
6. Vecteur recombinant selon l'une des revendications 1 à 5, caractérisé en ce que le terminateur de transcription
15 actif chez les mycobactéries est le terminateur du coliphage T4 (*tT4*).
7. Vecteur recombinant selon l'une des revendications 1 à 6, caractérisé en ce qu'il est un plasmide choisi parmi les
20 plasmides suivants déposés à la CNCM (Collection Nationale de Cultures de Microorganismes, Paris, France) :
- a) pJVEDa déposé à la CNCM sous le N° I-1797, le 12/12/1996,
- b) pJVEDb déposé à la CNCM sous le N° I-1906, le 25 juillet
25 1997,
- c) pJVEDc déposé à la CNCM sous le N° I-1799, le 12/12/1996.
8. Vecteur recombinant selon l'une des revendications 1 à
30 7, caractérisé en ce qu'il comprend en l'un des sites de clonage du polylinker une séquence d'acide nucléique de mycobactérie chez laquelle on détecte un polypeptide susceptible d'être exporté et/ou sécrété, et/ou d'être induit ou réprimé lors de l'infection par ladite
35 mycobactérie ou encore exprimé ou produit de façon constitutive, ainsi que les séquences promotrices et/ou régulatrices associées susceptibles de permettre ou de

favoriser l'exportation et/ou la sécrétion dudit polypeptide, ou tout ou partie de gène codant pour ledit polypeptide.

9. Vecteur recombinant selon l'une des revendications 1 à 8, caractérisé en ce que la séquence d'acide nucléique de mycobactérie qu'il contient est obtenue par fragmentation physique ou par digestion enzymatique de l'ADN génomique ou de l'ADN complémentaire d'un ARN d'une mycobactérie.

10

10. Vecteur recombinant selon l'une des revendications 1 à 9, caractérisé en ce que ladite mycobactérie est *M. tuberculosis*.

11. Vecteur recombinant selon l'une des revendications 1 à 9, caractérisé en ce que ladite mycobactérie est choisie parmi *M. africanum*, *M. bovis*, *M. avium* ou *M. leprae*.

12. Vecteur recombinant selon la revendication 10, caractérisé en ce qu'il est un plasmide choisi parmi les plasmides suivants déposés à la CNCM :

- a) p6D7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1814,
- b) p5A3 déposé le 28 janvier 1997 à la CNCM sous le N°I-1815,
- c) p5F6 déposé le 28 janvier 1997 à la CNCM sous le N°I-1816,
- d) p2A29 déposé le 28 janvier 1997 à la CNCM sous le N°I-1817,
- e) pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818,
- f) p5B5 déposé le 28 janvier 1997 à la CNCM sous le N°I-1819,
- g) p1C7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1820,
- h) p2D7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1821,

- i) p1B7 déposé le 31 janvier 1997 à la CNCM sous le N°I-1843,
j) pJVED/M. tuberculosis déposé le 25 juillet 1997 à la CNCM sous le N° I-1907,
5 k) pM1C25 déposé le 4 août 1998 à la CNCM sous le n°I-2062.

13. Vecteur recombinant selon la revendication 12, caractérisé en ce qu'il s'agit du plasmide pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818.

10

14. Procédé de criblage de séquences de nucléotides issues de mycobactéries pour déterminer la présence de séquences correspondant à des polypeptides exportés et/ou sécrétés pouvant être induits ou réprimés lors de l'infection, leurs
15 séquences promotrices et/ou régulatrices associées susceptibles notamment de permettre ou de favoriser l'exportation et/ou la sécrétion desdits polypeptides d'intérêt, ou tout ou partie de gènes d'intérêt codant pour lesdits polypeptides, caractérisé en ce qu'il met en oeuvre
20 un vecteur selon l'une des revendications 1 à 13.

15. Procédé de criblage selon la revendication 14, caractérisé en ce qu'il comprend les étapes suivantes :

- a) la fragmentation physique des séquences d'ADN de
25 mycobactéries ou leur digestion par au moins une enzyme déterminée et la récupération des fragments obtenus ;
b) l'insertion des fragments obtenus à l'étape a) dans un site de clonage, compatible le cas échéant avec l'enzyme de l'étape a), du polylinker d'un vecteur selon l'une des
30 revendications 1 à 13 ;
c) si besoin, l'amplification desdits fragments contenus dans le vecteur, par exemple par répllication de ce dernier après insertion du vecteur ainsi modifié dans une cellule déterminée, de préférence *E coli* ;
35 d) la transformation de cellules hôtes par le vecteur amplifié à l'étape c), ou en l'absence d'amplification, par le vecteur de l'étape b) ;

e) la culture de cellules hôtes transformées dans un milieu permettant la mise en évidence du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs contenu dans le vecteur ;

5 f) la détection des cellules hôtes positives (colonies positives) pour l'expression du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs ;

10 g) l'isolement de l'ADN des colonies positives et l'insertion de cet ADN dans une cellule identique à celle de l'étape c) ;

15 h) la sélection des insertions contenues dans le vecteur, permettant l'obtention de clones positifs pour le marqueur d'exportation et/ou de sécrétion, et /ou pour le marqueur d'activité de promoteurs ;

i) l'isolement et la caractérisation des fragments d'ADN de mycobactéries contenues dans ces insérats, et l'étape i) du procédé pouvant comporter en outre une étape de séquençage des insertions sélectionnées.

20

16. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie, caractérisée en ce qu'elle est obtenue par un procédé selon la revendication 14 et/ou un procédé comprenant les étapes a) et b), ou a), b) et c) du procédé
25 selon la revendication 15.

30

17. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie selon la revendication 16, caractérisée en ce que ladite mycobactérie est une mycobactérie pathogène.

18. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie selon la revendication 17, caractérisée en ce que ladite mycobactérie est une mycobactérie appartenant au groupe du complexe *Mycobacterium tuberculosis*.

19. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie selon la revendication 18, caractérisée en ce que ladite mycobactérie est *Mycobacterium tuberculosis*.
- 5 20. Séquence nucléotidique de mycobactérie ou comprenant une séquence nucléotidique de mycobactérie susceptible d'être sélectionnée par un procédé selon l'une des revendications 14 et 15.
- 10 21. Séquence nucléotidique de mycobactérie ou comprenant une séquence nucléotidique de mycobactérie selon la revendication 20, caractérisée en ce que ladite mycobactérie est choisie parmi *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. avium*, *M. leprae*, *M. paratuberculosis*, *M.*
15 *kansasii* ou *M. xenopi*.
22. Séquence nucléotidique selon l'une des revendications 20-21 caractérisée en ce qu'elle est choisie parmi les séquences de fragments d'ADN de mycobactérie de séquence
20 nucléique SEQ ID N°1 à SEQ ID N° 24C, SEQ ID N°27A à SEQ ID N°27C, SEQ ID N°29 et SEQ ID N° 31A à SEQ ID N°50F.
23. Séquence nucléotidique de mycobactérie l'une des revendications 20-21 caractérisée en ce qu'elle est choisie
25 parmi les séquences de fragments d'ADN de mycobactérie de séquence SEQ ID N°1, SEQ ID N°3A, SEQ ID N°5A, SEQ ID N°6A, SEQ ID N°7A, SEQ ID N°8A, SEQ ID N°9A, SEQ ID N°10A, SEQ ID N°27A ou SEQ ID N°29 contenus respectivement dans les vecteurs pDP428 (CNCM, N°I-1818), p6D7 (CNCM, N°I-1814),
30 p5F6 (CNCM, N°I-1816), p2A29 (CNCM, N°I-1817), p5B5 (CNCM, N°I-1819), p1C7 (CNCM, N°I-1820), p2D7 (CNCM, N°I-1821), p1B7 (CNCM, N°I-1843), p5A3 (CNCM, N°I-1815) et pM1C25 (CNCM, N°I-2062).
- 35 24. Séquence nucléotidique comprenant la totalité du cadre ouvert de lecture d'une séquence selon l'une quelconque des revendications 20 à 23.

25. Polynucléotide caractérisé en ce qu'il comprend un polynucléotide choisi parmi :

- a) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide selon l'une des revendications 20 à 24,
- b) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide selon l'une des revendications 20 à 24,
- c) un polynucléotide hybridant dans des conditions de forte stringence avec une séquence de polynucléotide selon l'une des revendications 20 à 24,
- d) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini selon l'une des revendications 20 à 24 ou défini en a).

26. Polypeptide, leurs fragments ou fragments biologiquement actifs ou leurs polypeptides homologues, susceptible d'être codé par une séquence nucléotidique de mycobactérie selon l'une des revendications 20 à 25, et susceptible d'être exporté et/ou sécrété, et/ou induit ou réprimé, ou exprimé de façon constitutive lors de l'infection.

27. Mycobactérie recombinante caractérisée en ce qu'elle est transformée par un vecteur recombinant selon l'une des revendications 1 à 13.

28. Polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques de séquence SEQ ID N°1 à SEQ ID N°2.

29. Polynucléotide caractérisé en ce qu'il comprend un polynucléotide choisi parmi :

- a) un polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques SEQ ID N°1 à SEQ ID N°2,

- b) un polynucléotide dont la séquence nucléique est la séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses, de la séquence SEQ ID N°1,
- 5 c) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide défini en a) ou b),
- d) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide défini en a), b) ou c),
- e) un polynucléotide hybridant dans des conditions de forte
- 10 stringence avec une séquence de polynucléotide défini en a), b), c) ou d),
- f) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini en a), b), c), d) ou e).
- 15 30. Polynucléotide selon l'une des revendications 20 à 25, 28 et 29, caractérisé en ce que sa séquence nucléique hybride avec l'ADN de séquence de mycobactéries et préférentiellement avec de l'ADN de séquences de mycobactéries appartenant au complexe de *Mycobacterium*
- 20 *tuberculosis*.
31. Polypeptide caractérisé en ce qu'il est codé par une séquence polynucléotidique selon l'une des revendications 20 à 25.
- 25 32. Polypeptide caractérisé en ce qu'il comprend un polypeptide choisi parmi :
- a) un polypeptide dont la séquence d'acides aminés est comprise dans une séquence d'acides aminés choisie parmi
- 30 les séquences d'acides aminés SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 28 et SEQ ID N° 30 à SEQ ID N° 50F,
- b) un polypeptide homologue au polypeptide défini en a),
- c) un fragment d'au moins 5 acides aminés d'un polypeptide
- 35 défini en a) ou b),
- d) un fragment biologiquement actif d'un polypeptide défini en a), b), ou c).

33. Polypeptide dont la séquence d'acides aminés est comprise dans la séquence d'acides aminés SEQ ID N°1 ou SEQ ID N°2, ou polypeptide de séquence d'acides aminés SEQ ID N°28.

34. Polynucléotide caractérisé en ce qu'il code pour un polypeptide selon l'un des revendications 32 et 33.

35. Séquence d'acide nucléique utilisable comme amorce, caractérisée en ce que ladite séquence est choisie parmi les séquences d'acide nucléique de polynucléotide selon l'une des revendications 20 à 25, 28 à 30, et 34.

36. Séquence d'acide nucléique selon la revendication 35, caractérisée en ce que ladite séquence est choisie parmi les séquences SEQ ID N°25 et SEQ ID N°26.

37. Séquence d'acide nucléique selon l'une des revendications 35 et 36 pour la détection et/ou l'amplification de séquences nucléiques.

38. Séquence d'acide nucléique utilisable comme sonde, caractérisée en ce que ladite séquence est choisie parmi les séquences d'acide nucléique selon l'une des revendications 20 à 25, 28 à 30, et 34.

39. Séquence d'acide nucléique selon la revendication 38, caractérisée en ce qu'elle est marquée par un composé radioactif ou par un composé non radioactif.

40. Séquence d'acide nucléique selon l'une des revendications 38 et 39, caractérisée en ce que celle-ci est immobilisée sur un support, de manière covalente ou non-covalente.

41. Séquence d'acide nucléique selon l'une des revendications 38 à 40 pour la détection et/ou l'amplification de séquences nucléiques.
- 5 42. Séquence d'acide nucléique selon l'une des revendications 38 à 41, caractérisée en ce que ladite séquence est une séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses, de la séquence SEQ ID N°1.
- 10 43. Vecteur recombinant de clonage, d'expression et/ou d'insertion, caractérisé en ce qu'il contient une séquence d'acide nucléique de polynucléotide selon l'une des revendications 20 à 25, 28 à 30, et 34.
- 15 44. Cellule hôte, caractérisée en ce qu'elle est transformée par un vecteur recombinant selon la revendication 43.
- 20 45. Cellule hôte selon la revendication 44, caractérisée en ce qu'il s'agit de la souche de *E. coli* transformée par le plasmide pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818 ou transformée par le plasmide pMIC25 déposé le 4 août 1998 à la CNCM sous le N°I-2062, ou d'une souche de
- 25 *M. tuberculosis*, *M. bovis* ou *M. africanum* possédant potentiellement tous les systèmes de régulation appropriés.
46. Procédé de préparation d'un polypeptide, caractérisé en ce qu'il met en oeuvre un vecteur selon la revendication
- 30 43.
47. Polypeptide recombinant susceptible d'être obtenu par un procédé selon la revendication 46.
- 35 48. Polypeptide hybride, caractérisé en ce qu'il comporte au moins la séquence d'un polypeptide selon l'une des revendications 26, 32, 33 et 47 et une séquence d'un

polypeptide susceptible d'induire une réponse immunitaire chez l'homme ou l'animal.

49. Polypeptide hybride selon la revendication 48, caractérisé en ce que le polypeptide susceptible d'induire une réponse immunitaire contient au moins un déterminant antigénique capable d'induire une réponse humorale et/ou cellulaire.

50. Polynucléotide codant pour un polypeptide hybride selon l'une des revendications 48 et 49.

51. Polypeptide hybride selon l'une des revendications 48 et 49, caractérisée en ce qu'il s'agit d'une protéine recombinante obtenue par l'expression d'un polynucléotide selon la revendication 50.

52. Procédé pour la détection *in vitro* d'anticorps dirigés contre une mycobactérie et préférentiellement une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

a) mise en contact de l'échantillon biologique avec un polypeptide selon l'une des revendications 26, 32, 33 et 47 ;

b) mise en évidence du complexe antigène-anticorps formé.

53. Procédé pour la détection d'une infection par une mycobactérie et préférentiellement une bactérie du complexe *Mycobacterium tuberculosis* dans un mammifère, caractérisé en ce qu'il comprend les étapes suivantes :

a) préparation d'un échantillon biologique contenant des cellules dudit mammifère plus particulièrement des cellules du système immunitaire dudit mammifère et plus particulièrement encore des cellules T ;

b) incubation de l'échantillon biologique de l'étape a) avec un polypeptide selon l'une quelconque des revendications 26, 32, 33 et 47 ;

c) détection d'une réaction cellulaire indiquant une sensibilisation préalable du mammifère audit polypeptide notamment la prolifération cellulaire et/ou la synthèse de protéines telles que l'interféron gamma;

d) détection d'une réaction d'hypersensibilité retardée ou de sensibilisation du mammifère audit polypeptide.

54. Kit pour le diagnostic in vitro d'une infection par une mycobactérie appartenant au complexe *Mycobacterium tuberculosis*, comprenant :

a) un polypeptide selon l'une des revendications 26, 32, 33 et 47 ;

b) le cas échéant, les réactifs pour la constitution du milieu propice à la réaction immunologique;

c) les réactifs permettant la détection des complexes antigène-anticorps produits par la réaction immunologique ;

d) le cas échéant, un échantillon biologique de référence (témoin négatif) dépourvu d'anticorps reconnus par ledit polypeptide ;

e) le cas échéant, un échantillon biologique de référence (témoin positif) contenant une quantité prédéterminée d'anticorps reconnus par ledit polypeptide.

55. Anticorps mono- ou polyclonaux, leurs fragments, ou anticorps chimériques, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une des revendications 26, 32, 33 et 47 .

56. Anticorps selon la revendication 55, caractérisé en ce qu'il s'agit d'un anticorps marqué.

57. Procédé pour la détection spécifique de la présence d'un antigène d'une bactérie du complexe *Mycobacterium*

tuberculosis dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact de l'échantillon biologique avec un anticorps selon l'une des revendications 55 et 56 ;
- 5 b) mise en évidence du complexe antigène-anticorps formé.

58. Kit pour la détection spécifique de la présence d'un antigène d'une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

- 10 a) un anticorps polyclonal ou monoclonal selon l'une des revendications 55 et 56 ;
- b) les réactifs pour la constitution du milieu propice à la réaction immunologique ;
- 15 c) les réactifs permettant la détection des complexes antigène-anticorps produits par la réaction immunologique.

59. Procédé de détection et d'identification rapide d'une mycobactérie et préférentiellement de *M. tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comporte les étapes suivantes :

- a) isolement de l'ADN à partir de l'échantillon biologique à analyser, ou obtention d'un ADNc à partir de l'ARN de l'échantillon biologique ;
- 25 b) amplification spécifique de l'ADN des mycobactéries appartenant au complexe *Mycobacterium tuberculosis* à l'aide d'amorces selon l'une des revendications 35 à 37 ;
- c) analyse des produits d'amplification.

30 60. Procédé pour la détection de bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact d'une sonde oligonucléotidique selon l'une des revendications 38 à 42 avec un échantillon biologique, l'ADN contenu dans l'échantillon biologique
- 35

ayant, le cas échéant, préalablement été rendu accessible à l'hybridation, dans des conditions permettant l'hybridation de la sonde à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;

- 5 b) détection de l'hybride formé entre la sonde oligonucléotidique et l'ADN de l'échantillon biologique.

61. Procédé pour la détection de bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon
10 biologique, caractérisé en ce qu'il comprend les étapes suivantes :

a) mise en contact d'une sonde oligonucléotidique immobilisée sur un support selon la revendication 40 avec un échantillon biologique, l'ADN de l'échantillon, ayant,
15 le cas échéant, été préalablement rendu accessible à l'hybridation, dans des conditions permettant l'hybridation de la sonde à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;

b) mise en contact de l'hybride formé entre la sonde
20 oligonucléotidique immobilisée sur un support et l'ADN contenu dans l'échantillon biologique, le cas échéant après élimination de l'ADN de l'échantillon biologique n'ayant pas hybridé avec la sonde, avec une sonde oligonucléotidique marquée selon la revendication 39.

25 62. Procédé de détection selon la revendication 61, caractérisé en ce que, préalablement à l'étape a), l'ADN de l'échantillon biologique, ou l'ADNc obtenu par transcription inverse de l'ARN de l'échantillon, est
30 amplifié à l'aide d'un couple d'amorces selon l'une des revendications 35 à 37.

63. Procédé pour la détection de la présence d'une bactérie appartenant au complexe de *Mycobacterium tuberculosis* dans
35 un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact de l'échantillon biologique avec un couple d'amorces selon l'une des revendications 35 à 37, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation des amorces à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;
- b) amplification de l'ADN de la bactérie du complexe *Mycobacterium tuberculosis* ;
- c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par les amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique marquée selon la revendication 39.

64. Procédé pour la détection de la présence d'une bactérie appartenant au complexe de *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact de l'échantillon biologique avec deux couples d'amorces selon l'une des revendications 35 à 37, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation des amorces à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;
- b) amplification de l'ADN de la bactérie du complexe *Mycobacterium tuberculosis* ;
- c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par lesdites amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique marquée selon la revendication 39.

65. Kit pour la détection de la présence d'une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

- a) une sonde oligonucléotidique selon l'une des revendications 38 à 42 ;
- b) les réactifs nécessaires à la mise en oeuvre d'une réaction d'hybridation ;
- 5 c) le cas échéant, un couple d'amorces selon l'une des revendications 35 à 37 ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN (ADN génomique, plasmidique ou ADNc) d'une bactérie du complexe *Mycobacterium tuberculosis*.

10

66. Kit ou nécessaire pour la détection de la présence d'une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

15

- a) une sonde oligonucléotidique, dite sonde de capture, selon la revendication 40 ;
 - b) une sonde oligonucléotidique, dite sonde de révélation, selon l'une des revendications 38 à 42 ;
 - c) le cas échéant, un couple d'amorces selon l'une des
- 20 revendications 35 à 37 ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis*.

25

67. Kit pour l'amplification de l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* présent dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

30

- a) un couple d'amorces selon l'une des revendications 35 à 37 ;
 - b) les réactifs nécessaires pour effectuer une réaction d'amplification d'ADN ;
 - c) éventuellement un composant permettant de vérifier la séquence du fragment amplifié, plus particulièrement une sonde oligonucléotidique selon l'une des revendications 38
- 35 à 42.

[2]

68. Composition immunogène caractérisée en ce qu'elle comprend un ou plusieurs polypeptides selon l'une des revendications 26, 32, 33 et 47 et/ou un ou plusieurs polypeptides hybrides selon l'une des revendications 48, 49 et 51.

69. Vaccin caractérisé en ce qu'il contient un ou plusieurs polypeptides selon l'une des revendications 26, 32, 33 et 47 et/ou un ou plusieurs polypeptides hybrides selon l'une des revendications 48, 49 et 51, en association avec un véhicule pharmaceutiquement compatible et, le cas échéant un ou plusieurs adjuvants de l'immunité appropriés.

70. Vaccin destiné à l'immunisation à l'encontre d'une infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, comprenant un vecteur selon la revendication 43 ou un polynucléotide selon la revendication 50, en association avec un véhicule pharmaceutiquement acceptable.

71. Vaccin caractérisé en ce qu'il contient une ou plusieurs séquences nucléotidiques selon l'une des revendications 20 à 24 et/ou un ou plusieurs polynucléotides selon la revendication 25 en association avec un véhicule pharmaceutiquement compatible et, le cas échéant, un ou plusieurs adjuvants de l'immunité appropriés.

72. Méthode de criblage de molécules capables d'inhiber la croissance de mycobactéries ou le maintien de mycobactéries dans un hôte, caractérisée en ce que lesdites molécules bloquent la synthèse ou la fonction des polypeptides codés par une séquence nucléotidique selon l'une quelconque des revendications 20 à 24 ou par un polynucléotide selon la revendication 25.

73. Méthode de criblage selon la revendication 72, caractérisée en ce que les molécules sont des anti-messagers ou induisent la synthèse d'anti-messagers.

5 74. Molécules capables d'inhiber la croissance de mycobactéries ou le maintien de mycobactéries dans un hôte, caractérisées en ce que lesdites molécules sont synthétisées d'après la structure des polypeptides codés par une séquence nucléotidique selon l'une quelconque des
10 revendications 20 à 24 ou par un polynucléotide selon la revendication 25.

15

20

1/185

```

1  GGATCCGAGGGAACCTGACC ATG GTC GTA GGG ATG ACT TGA CAGTTTCACGCGGGTCCGACGACCGTTCCCG 72
1      M V V G M T *
73 TCAGAGGCGATACGTTGGTGGACACGCTCGGAAGCTGGGAGGTGAATCTG ATG GCT GGC GAC CAA GAG CTG 144
1      M A G D Q E L
145 GAA CTG CGG TTC GAC GTT CCT CTT TAC ACC CTT GGC GAG GCA TCG CGG TAC CTG GTG GTT 204
8 E L R P D V P L Y T L A E A S R Y L V V
205 CGC CGC GCC ACC CTG GCT ACC TGG GCT GAC GGC TAC GAG CGT CGG CGC GGC AAC GCA CCG 264
28 P R A T L R T W A D G Y E R R P A N A P
365 GGG GTC CAG GGG CAA CGG ATC GGC TTT GAC GGC TAT TCG CTC GCG CAG CTT TTT GGC GAC 324
48 A V Q G Q P I A F D A Y S V A Q L P G D
67
325 GTC ACT GGT GCG GCG GTT GCG GGC GTC CAG CGC CAG CGA CAC CAC ATA CCG CGG GTC CCG 384
68 V T G A R V A G V Q P Q R H H I R P V R
87
385 TTG CGG GGG CGG TTG GGT GGG GTT GGG TCG CTC GGT CAC GCG AGG CAG TTC GGT GGC TAT 444
80 L R G P L G G V G C L R H P R Q F A G Y
107
445 TTG TCG CAG TAG CCGAGCGGCAATGTCTG ATG TCT TGG TAG CTACGATCGGCTCGGGGGGCGGTACCGCG 515
100 L S Q * M S W *
516 CCGAGCGCGCGCTCTCCGCTGCGGCTGAGTGGTGGTGGACGACCA ATG ACT GCG ACC CGG 587
1      M T A T R
588 CGA CTT CGA AAC CGC CAC CGG TTA GAT TCC CGG ACT GCG TCA TCG GCA GGT AAA CGG CGG 647
6 R L R N R R R L D S P T A S S P G K P P
725
648 GCA CTA ACC CCA GCA ACC AAC CGG TGA AGAGCAACCAACGCGACCTCCGCGAGTTCCGCGCTTAACCGCATC 718
26 A L T P A T N P *
719 ATG AAC TCG TCG ATT TCG GAC TCC CGG TAC TCT CGC GCA GTG CGT GCG CGC CAG GGT ACC 778
1 M N C W I S D S P Y S R A V R A R R P T
80
779 GAA GAT CGC GTC CAT GCG TTC GGC GTG GAC CGC ACA GCA CCG GGA GTT GUC CGC GCG GAG 838
21 E D R V H A P G V D R T A P C V G G A E
90
839 GGC CGA GAT GGC AGG ATG ACC GAT CGT CGG GCG CGG GAA CTC CCA GCG CGC CGG ACC GTC 898
41 G R D G R M T D R R G R E L P G R R T V
100
899 GCA AAC CCG TCG CAA ACC CGT CGC AAA CGG TAA GCGATCATCC ATG AAG ACA GCG ACC CGG 958
61 A N P S Q T R R K P * M N T G T A
960 ACC ACC CGG CGC ACC CTG TTG GCA GTA CTG ATC GCG CTC GCG TTG CGG GGG GCG GCG GTT 1018
7 T T R R R L L A V L I A L A L P G A A V
1020 GCG CTG CTG GCG GAA CCA TCA GCG ACC GCG GCG TCG CAG CGG TCG GCG GCG ACC GAA GTG 1078
27 R L L A E P S A T G A S D P C A A S E V
1080 GCG AGG ACG GTC GGT TCG GTC GCG AAG TCG ATG GCG CAC TAC CTG GAT TCA CAC GCA GAG 1138
47 R R T V G S V A K S M G D Y L D S H P E
1140 ACC AAC CAG GTC ATG ACC GCG GTC TTG CAG CAG CAG GTA GCG CGG GCG TCG CTC GCA TCG 1198
67 T N Q V M T A V L Q Q Q V G P Q S V A S
1200 CTG AAG GCG CAG TTC GAG GCG AAT GCG AAG GTC GCA TCG GAT CC 1240
87 L K A N P E A N S K V A S D
100

```

SEQ ID N° 1

FIGURE 1

FEUILLE DE REMPLACEMENT (REGLE 26)

2/185

Insert du clone contenant BP428 et contenu dans seq1

```

1/1
31/11
GAT CGC CTT TGA CGC CTA TTC GGT CGC GCA GCT TTT TGG CGA CGT CAC TGG TGC CCG CGT
asp arg leu GPA arg leu phe gly arg ala ala phe trp arg arg his trp cys pro arg
61/21
91/31
TGC GGG CGT CCA GCG GCA GCG ACA CCA CAT ACG GGC GGT CCG GTT GCG GCG GCG GTT GCG
cys gly arg pro ala ala ala thr pro his thr ala gly pro val ala gly ala val gly
121/41
151/51
TGG GGT TGG GTG CGT CCG TCA CCC CAG GCA GTT CCG TGG CTA TTT GTC GCA GTA GCG CCA
trp gly trp val pro pro ser pro gln ala val arg trp leu phe val ala val ala arg
181/61
211/71
CGG CAT TGT CGA TGT CTT GGT AGC TAG CAT CCG GTC GGG GGG CCG CTA CCA GCG CCA GCG
arg his cys arg cys leu gly ser AMB his pro val gly gly pro leu pro ala pro ala
241/81
271/91
CCG GGG CTC CCG GGT CCG GGT AGT GCG CGT CGA GTT GGT CGT GCA CCA GCA ATG ACT GCG
pro gly leu pro gly pro gly ser ala arg arg val gly arg gly pro ala met thr ala
301/101
331/111
ACG CCG CGA CTT CGA AAC CCG CAC CCG TTA CAT TCC CCG ACT GCG TCA TCG CCA GGT AAA
thr arg arg leu arg asn arg his arg leu asp ser pro thr ala ser ser pro gly lys
361/121
391/131
CCG CCG GCA CTA ACG CCA GCA ACC AAC CCG TGA AGA CCA ACC AAC GCG ACC TCG CCA GGT
pro pro ala leu thr pro ala thr asn pro GPA arg pro thr asn gly thr cys ala gly
421/141
451/151
TGC GCG TCA ACC GCA TCA TGA ACT GGT GCA TTT CCG ACT CCC CGT ACT CTC GCG CAG TGC
cys gly ser thr ala ser GPA thr ala gly phe asg thr pro arg thr leu ala gln cys
481/161
511/171
GTG CCC GCG ACC CTA CCG AAG ATC CCG TGC ATG CGT TCG GCG TGG ACC GCA CAG CAC CTG
val pro ala ser leu pro lys ile ala cys met arg ser ala trp thr ala gln his leu
541/181
571/191
GAG TTT GCG GCG CCG AGG CCG GAG ATG GCA GGA TGA CCG ATC GTC GCG GCG GCG AAC TCC
gln leu ala ala pro arg ala glu met ala gly GPA arg ile val gly gly gly asn ser
601/201
631/211
CAG GCC GCG GGA CCG TCG CAA ACC CGT CCG AAA CCC GTC GCA AAC CGT AAG GAG TCA TCC
gln ala ala gly pro ser gln thr arg arg lys pro val ala asn arg lys glu ser ser
661/221
691/231
ATG AAG ACA GCG ACC GCG ACG ACG GCG CCG ACG CTG TTG GCA GTA CTG ATC GCG CTC GCG
met lys thr gly thr ala thr thr arg arg arg leu leu ala val leu ile ala leu ala
721/241
751/251
TTG CCG GGG GCG GCG GTT GCG CTG CTG GCG GAA CCA TCA GCG ACC GCG GCG TCG GAC CCG
leu pro gly ala ala val ala leu leu ala glu pro ser ala thr gly ala ser asp pro
781/261
811/271
TGC CCG GCG ACC GAA GTG GCG ACG ACG GTC GGT TCG GTC CCG AAG TCG ATG GCG GAG TAC
cys ala ala ser glu val ala arg thr val gly ser val ala lys ser met gly asp tyr
841/281
871/291
CTG GAT TCA CAC CCA GAG ACC AAC CAG GTG ATG ACC GCG GTC TTG CAG CAG CAG GTA GCG
leu asp ser his pro glu thr asn gln val met thr ala val leu gln gln gln val gly
901/301
931/311
CCG GGG TCG GTC GCA TCG CTG AAG GCG CAT TTC GAG GCG AAT CCG AAG GTC GCA TCG GAT C
pro gly ser val ala ser leu lys ala his phe glu ala asn pro lys val ala ser asp

```

SEQ ID N° 1A'

FIGURE 1A'

3/185

Insert du clone contenant DP428, autre phase de lecture

2/1 32/11
 ATC GCC TTT GAC GCC TAT TCG GTC GCG CAG CTT TTT GGC GAC GTC ACT GGT GCC CGC GTT
 ile ala phe asp ala tyr ser val ala gln leu phe gly asp val thr gly ala arg val
 62/21 82/31
 GCG GGC GTC CAG CCG CAG CGA CAC CAC ATA CCG CCG GTC CCG TTG CCG GGG CCG TTG GGT
 ala gly val gln pro gln arg his his ile arg pro val arg leu arg gly pro leu gly
 122/41 152/51
 GGG GTT GGG TGC CTC CGT CAC CCC AGG CAG TTC GCT GGC TAT TTG TCG CAG TAG CGC GAC
 gly val gly cys leu arg his pro arg gln phe ala gly tyr leu ser gln AMB arg asp
 182/61 212/71
 GGC ATT GTC GAT GTC TTG GTA GCT AGC ATC CCG TCG GGG GGC CGC TAC CAG CGC CAG CGC
 gly ile val asp val leu val ala ser ile arg ser gly gly arg tyr gln arg gln arg
 242/81 272/91
 CCG GGC TCC CCG GTC CCG GTA GTG CCG GTC CAG TTG GTC GTG GAC CAG CAA TGA CTC CGA
 arg gly ser pro val arg val val arg val gln leu val val asp gln gln GFA leu arg
 302/101 332/111
 CCC GGC GAC TTC GAA ACC GCC ACC GGT TAG ATT CCC CGA CTG CGT CAT CGC CAG GTA AAC
 pro gly asp phe gln thr ala thr gly AMB ile pro arg leu arg his arg gln val ser
 362/121 392/131
 CSC CGG CAC TAA CGC CAG CAA CCA ACC CGT GAA GAC CAA CCA ACG GCA CCT GCG CAG GTT
 arg arg his OCH arg gln gln pro thr arg gln asp gln pro thr ala pro ala gln val
 422/141 452/151
 GCG GCT CAA CCG CAT CAT GAA CTG CTG GAT TTC GGA CTC CCC GTA CTC TCG CGC AGT GCG
 ala ala gln pro his his glu leu leu asp phe gly leu pro val leu ser arg ser ala
 482/161 512/171
 TCG CCG CGA GCC TAC CGA AGA TCG CGT GCA TGC GTT CCG CGT GGA CCG CAC AGC ACC TGG
 cys pro arg ala tyr arg arg ser arg ala cys val arg arg gly pro his ser thr trp
 542/181 572/191
 AGT TGG CCG CGC CGA GGG CCG AGA TGG CAG GAT CAC GGA TCG TCG GGG GCG GGA ACT CCC
 ser trp arg arg arg gly pro arg trp gln asp asp gly ser ser gly ala gly thr pro
 602/201 632/211
 AGG CCG CCG GAC CGT CGC AAA CCC GTC GCA AAC CCG TCG CAA ACC GTA AGG AGT CAT CCA
 arg pro pro asp arg arg lys pro val ala asn pro ser gln thr val arg ser his pro
 662/221 692/231
 TGA AGA CAG GCA CCG CGA CGA CGC GGC GCA GGC TGT TGG CAG TAC TGA TCG CCC TCG CGT
 GFA arg gln ala pro arg arg arg gly ala gly cys trp gln tyr GFA ser pro ser arg
 722/241 752/251
 TGC CCG GGG CCG CCG TTG CGC TGC TGG CCG AAC CAT CAG CGA CCG GCG CGT CCG ACC CGT
 cys arg gly pro pro leu arg cys trp pro asn his gln arg pro ala arg arg thr arg
 782/261 812/271
 GCG CCG CCA GCG AAG TCG CGA GGA CCG TCG GTT CCG TCG CCA AGT CGA TGG GCG AGT ACC
 ala arg pro ala lys trp arg gly arg ser val arg ser pro ser arg trp ala thr thr
 842/281 872/291
 TGG AAT CAC ACC CAG AGA CCA ACC AGG TGA TGA CCG CCG TCT TGC AGC AGC AGG TAG GCG
 trp ile his thr gln arg pro thr arg GFA GFA pro arg ser cys ser ser arg AMB gly
 902/301 932/311
 CCG GGT CCG TCG CAT CGC TGA AGG CCC ATT TCG AGG CGA ATC CCA AGG TCG CAT CCG ATC
 arg gly arg ser his arg GFA arg pro ile ser arg arg ile pro arg ser his arg ile

SEQ ID N° 1B'

FIGURE 1B'

4/185

Seq1C: Insert du clone D9428, autre phase de lecture

```

3/1
TCG CCT TTG ACG CCT ATT CCG TCG CSC ACC TTT TTG GCG ACC TCA CTG GTG CCC GCG TTG
ser pro leu thr pro ile arg ser arg ser phe leu ala thr ser leu val pro ala leu
63/21
CGG GCG TCC AGC CCG AGC GAC ACC ACA TAC GCG CCG TCC GGT TCG GGG GGC COT TGG GTG
arg ala ser ser arg ser asp thr thr tyr gly arg ser gly cys gly gly arg trp val
123/41
GGG TTG GGT GGC TCC GTC ACC CCA GGC AGT TCG CTG GCT ATT TGT CCG AGT AGC GCG ACC
gly leu gly ala ser val thr pro gly ser ser leu ala ile cys arg ser ser ala thr
183/61
GCA TTG TCG ATG TCT TGG TAG CTA GCA TCC GGT CCG GCG GCG GGT ACC AGC GCG AGC GCG
ala leu ser met ser trp AMB leu ala ser gly arg gly ala ala thr ser ala ser ala
243/81
GGG GCT CCC CCG TCC GCG TAG TCG GCG TCG AGT TGG TCG TCG ACC AGC AAT GAC TCG GAC
gly ala pro arg ser gly AMS cys ala ser ser trp ser trp thr ser asn asp cys asp
303/101
CCG GCG ACT TCG AAA CCG CCA CCG GTT AGA TTC CCG GAC TCG GTC ATC GCG AGG TAA ACC
pro ala thr ser lys pro pro pro val arg phe pro asp cys val ile ala arg OCH thr
363/121
GCC GCG ACT AAC GCG ACC AAC CAA CCC GTG AAG ACC AAC CAA CCG CAC CTG CSC AGG TTG
ala gly thr asn ala ser asn gln pro val lys thr asn gla arg his leu arg arg leu
423/141
CGG CTC AAC CSC ATC ATG AAC TCG TCG ATT TCG GAC TCC CCG TAC TCT CSC GCA GTG COT
arg leu asn arg ile met asn cys trp ile ser asp ser pro tyr ser arg ala val arg
483/161
GCC CCG GAG CCT ACC GAA GAT CGC GTG CAT CCG TTC GCG GTG GAC CSC ACA GCA CCT GGA
ala arg gla pro thr glu asp arg val his ala phe gly val asp arg thr ala pro gly
543/181
GTT GCG GGC GTC GAG GCG CGA GAT GGC AGG ATC ACG GAT COT CCG GCG CCG GAA CTC CCA
val gly gly ala glu gly arg asp gly arg met thr asp arg arg gly arg glu leu pro
603/201
GGC CCG CCG ACC GTC GCA AAC CCG TCG CAA ACC COT CSC AAA CCG TAA GGA GTC ATC CAT
gly arg arg thr val ala asn pro ser gln thr arg arg lys pro OCH gly val ile his
663/221
GAA GAC AGG CAC CCG GAC GAC GCG GCG CAG GCT GTT GCG AGT ACT GAT CCG CCT CSC GTT
glu asp arg his arg asp asp ala ala gln ala val gly ser thr asp arg pro arg val
723/241
GCC GCG GGC CSC COT TCG GCT GCT GCG CGA ACC ATC AGC GAC CCG CCG GTC GGA CCC GTG
ala gly gly arg arg cys ala ala gly arg thr ile ser asp arg arg val gly pro val
783/261
CGC GCG CAG CGA AGT GCG GAG GAC GGT CCG TTC GGT CCG CAA GTC GAT GCG CGA CTA COT
arg gly gla arg ser gly glu asp gly arg phe gly arg gln val asp gly arg leu pro
843/281
GGA TTC ACA CCC AGA GAC CAA CCA GGT GAT GAC CCG GGT CTT GCA GCA GCA GGT AGG GCG
gly phe thr pro arg asp gln pro gly asp asp arg gly leu ala ala ala gly arg ala
903/301
GGG GTC GGT CCG ATC GCT GAA GGC CCA TTT CGA GCG GAA TCC CAA GGT CSC ATC GGA TC
gly val gly arg ile ala glu gly pro phe arg gly glu ser gln gly arg ile gly

```

SEQ ID N° 1C'

FIGURE 1C'

5/185

Séquence codante DP428 identique à la séquence Rv0203 prédite par Cole et al.
(Nature 393:537-544)

```

1/1                               31/11
ATG AAG ACA GGC ACC GCG ACC ACC CGG CGC AGG CTG TTG GCA GTA CTG ATC GCG CTC GCG
Met lys thr gly thr ala thr thr arg arg arg leu leu ala val leu ile ala leu ala
61/21                               91/31
TTG CCG GGG GCG GCC GTT GCG CTG CTG GCC GAA CCA TCA GCG ACC GCG GCG TCG GAC CCG
leu pro gly ala ala val ala leu leu ala glu pro ser ala thr gly ala ser asp pro
121/41                               151/51
TGC GCG GCC AGC GAA GTG GCG ACC ACC GTC GGT TCG GTC GCC AAG TCG ATC GCG GAC TAC
cys ala ala ser glu val ala arg thr val gly ser val ala lys ser met gly asp tyr
181/61                               211/71
CTG GAT TCA CAC CCA GAG ACC AAC CAG GTG ATG ACC GCG GTC TTG CAG CAG CAG GTA GCG
leu asp ser his pro glu thr asn gln val met thr ala val leu gln gln gln val gly
241/81                               271/91
CCG GGG TCG GTC GCA TCG CTG AAG GCC CAT TTC GAG GCG AAT CCC AAG GTC GCA TCG GAT
pro gly ser val ala ser leu lys ala his phe glu ala asn pro lys val ala ser asp
301/101                               331/111
CTG CAC GCG CTT TCG CAA CCG CTG ACC GAT CTT TCG ACT CCG TCG TCG CTG CCG ATC ACC
leu his ala leu ser gln pro leu thr asp leu ser thr arg cys ser leu pro ile ser
361/121                               391/131
GGC CTG CAG GCG ATC GGT TTG ATC CAG GCG GTG CAG GCG GCG CCG CCG TAG
gly leu gln ala ile gly leu met gln ala val gln gly ala arg arg AMB

```

SEQ ID N° 1D

FIGURE 1D

ORF contenant la séquence DP428 et faisant partie de seq1A'

```

1/1                               31/11
TGA CCG ATC GTC GCG GGC GGG AAC TCG CAG GCC GCG GCA CCG TCG CAA ACC CGT CCG AAA
GPA arg ile val gly gly gly asn ser gln ala ala gly pro ser gln thr arg arg lys
61/21                               91/31
GCC GTC GCA AAC CGT AAG GAG TCA TCC ATG AAG ACA GCG ACC GCG ACC ACC CCG CCG ACC
pro val ala asn arg lys glu ser ser met lys thr gly thr ala thr thr arg arg arg
121/41                               151/51
CTG TTG GCA GTA CTG ATC GCG CTC GCG TTG CCG GCG GCG GCG GTT GCG CTG CTG GCG GAA
leu leu ala val leu ile ala leu ala leu pro gly ala ala val ala leu leu ala glu
181/61                               211/71
CCA TCA GCG ACC GCG GCG TCG GAC CCG TCC GCG GCG ACC GAA CTG GCG ACG ACC GTC GGT
pro ser ala thr gly ala ser asp pro cys ala ala ser glu val ala arg thr val gly
241/81                               271/91
TCG CTC GCG AAG TCG ATG GCG GAC TAC CTG GAT TCA CAC CCA GAG ACC AAC CAG GTG ATC
ser val ala lys ser met gly asp tyr leu asp ser his pro gln thr asn gln val met
301/101                               331/111
ACC GCG CTC TTG CAG CAG CAG GTA GCG CCG GCG TCG GTC GCA TCG CTG AAG GCG CAT TTC
thr ala val leu gln gln gln val gly pro gly ser val ala ser leu lys ala his phe
361/121                               391/131
GAG GCG AAT CCC AAG GTC GCA TCG GAT CTG CAC GCG CTT TCG CAA CCG CTG ACC GAT CTT
glu ala asn pro lys val ala ser asp leu his ala leu ser gln pro leu thr asp leu
421/141                               451/151
TCG ACT GCG TCG TCG CTG CCG ATC ACC GCG CTG CAG GCG ATC GGT TTG ATG CAG GCG CTG
ser thr arg cys ser leu pro ile ser gly leu gln ala ile gly leu met gln ala val
481/161
CAG GCG GCG GCG CCG TAG
gln gly ala arg arg AMB

```

SEQ ID N° 1F

6/185

```

481 CCGCTCGGGGCGGCGCTACCGAGCGCAGCGCGCGGGCTCCCGGCTCGGGTA GTG CGC CTC GAG TTG CTC GTG 563
1 V R V C L V V 7
584 GAG CAG CAA TGA CTGCGACCGGGTACTTGGAAACCGCGCAGCTGTGATTCGCGACTGCTGCTGTCGCGAGTAA 639
8 D Q Q *
586 ACCGCGCGCGACTAGCGCGAGCGACGACGTC GTG AAG AGC AAC CAA CGG CAC CTG CGC AGG TTG CGG 705
1 V K T H Q R H L R H L R 12
706 CTC AAC CGC ATC ATG AAC TCG TGG ATT TCG GAC TCG CGC TAC TCT CGC GCA GTG GGT GGC 768
13 L H R I H N C N I S D S P Y S R A V R A 30
766 CGC GAG CCG ACC GAA GAT CGC GTG CAG GCG TTC GGC GTG GAC GCG ACA GCA GGT GGA GTT 825
33 R C F T S D R V H A F G V D R T A P G V 50
826 GCG CGC CGC GAG GCG CCA GAT GCG AGG ATG ACG GAT GGT CGG GAG CGG GAA CTC CCA GCG 885
53 G G A E G R D G R M T D R R G R E L F G 72
886 GCG CGC ACC GTC GCA AAC CGG TCG GAA ACC GGT CGC AAA CGG TAA GAGTCATCG ATG AAG 945
73 R R T V A H P S Q T R R K P * XXXXXX H E 2
947 ACA GCG ACC GCG AGC ACC CGC CGC AGG CTG TGG GGA GTA CTG ATC GCG CTC GCG TTG CGC 1004
3 T G T A T T T R R R L L A V L T S L A L P 32
1007 GCG GCG GCG GTT GCG CTG CTG GCG GAA CCA TCA GCG ACC GCG CGC TCG GAC CGC TGC GCG 1064
33 G A A V A L L A E F S A T G A S D P C A 42
1067 GCG AGC CAA GTG GCG AGG ACG CTC GGT TCG GTC GCG AAG TCG ATG GCG GAC TAC CTG GAT 1126
43 R S E V A R T V G S V A K S M G D Y L S 60
1127 TCA CAG CGA GAG ACC AAC CAG GTG ATG ACC GCG CTC TTG CAG CAG CAG GTA GAG CGC GCG 1184
63 S R F S T H Q V M T A V L Q Q Q V G F Q 82
1187 TCG CTC GCA TCG CTG AAG GCG GAT TTC GAG GCG AAT CGC AAG CTC GCA TCG GAT CTG CAC 1246
83 S V A S L K A M F E A H P K V A S D L H 100
1247 GCG GTT TCG CAA CGC CTG ACC GAT GTT TCG ACT CGG TCG TCG CTG GCG ATC AGC GCG CTG 1304
103 A L S Q P L T D L S T R C S C P T E S L 120
1307 CAG GCG ATC GGT TTG ATG CAG GCG GCG CAG GCG GCG GCG TGG ATG CGG GAC CGC CGC 1365
123 Q A L Q L M Q A V Q Q A A R * H P Q R R S 5
1367 CGC CTC CGC CGC AGT CCA GGT GAG GCA GCG CTC GCG TAC CGG GCG GGT CTC TCG CGC GGT 1426
6 R V R R S R R E A A V A Y R G D V S P P 25
1427 TGT GGT CGC AGG TCA GCG CTC GCG GGT GGA GGT TCG GGT GTG GTT TCG ACC GCG TCG TCG 1484
26 S G R R S G V G A G F C G V V S T G S S 45
1487 CAG GGT GTG CGC TCG GGT TCG ATG ACA AAT CGC AGG TTG GGA TCG GTT GCG GCG TCG CAA 1546
46 Q G V F C G H M T S R R F G S V G G S R 65
1547 TCG TTG T 1553
84 H L 67

```

SEQ ID N° 2

FIGURE 2